



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>C12N 15/31, C07K 14/35, A61K 38/16, C12N 15/62, G01N 33/569, C12Q 1/68, C12N 5/10, 1/21 // A61K 39/04, (C12N 1/21, 1:19)</b>		A2	(11) International Publication Number: <b>WO 97/09428</b>  (43) International Publication Date: <b>13 March 1997 (13.03.97)</b>																																								
(21) International Application Number: <b>PCT/US96/14674</b>  (22) International Filing Date: <b>30 August 1996 (30.08.96)</b>		(74) Agents: MAKI, David, J. et al.; Seed and Berry L.L.P., 6300 Columbia Center, 701 Fifth Avenue, Seattle, WA 98104-7092 (US).																																									
(30) Priority Data: <table border="0"><tr><td>08/523,436</td><td>1 September 1995 (01.09.95)</td><td>US</td></tr><tr><td>08/533,634</td><td>22 September 1995 (22.09.95)</td><td>US</td></tr><tr><td>08/620,874</td><td>22 March 1996 (22.03.96)</td><td>US</td></tr><tr><td>08/659,683</td><td>5 June 1996 (05.06.96)</td><td>US</td></tr><tr><td>08/680,574</td><td>12 July 1996 (12.07.96)</td><td>US</td></tr></table>		08/523,436	1 September 1995 (01.09.95)	US	08/533,634	22 September 1995 (22.09.95)	US	08/620,874	22 March 1996 (22.03.96)	US	08/659,683	5 June 1996 (05.06.96)	US	08/680,574	12 July 1996 (12.07.96)	US	(81) Designated States: AL, AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).																										
08/523,436	1 September 1995 (01.09.95)	US																																									
08/533,634	22 September 1995 (22.09.95)	US																																									
08/620,874	22 March 1996 (22.03.96)	US																																									
08/659,683	5 June 1996 (05.06.96)	US																																									
08/680,574	12 July 1996 (12.07.96)	US																																									
(71) Applicant: CORIXA CORPORATION [US/US]; Suite 464, 1124 Columbia Street, Seattle, WA 98104 (US).		Published <i>Without international search report and to be republished upon receipt of that report.</i>																																									
(72) Inventors: REED, Steven, G.; 2843 - 122nd Pine Place N.E., Bellevue, WA 98005 (US). <u>SKEIKY, Yasir, A., W.</u> ; 8327 - 25th Avenue N.W., Seattle, WA 98117 (US). DILLON, Davin, C.; 21607 N.E. 24th Street, Redmond, WA 98053 (US). CAMPOS-NETO, Antonio; 9308 N.E. Midship Court, Bainbridge Island, WA 98110 (US). HOUGHTON, Raymond; 2636 - 242nd Place S.E., Bothell, WA 98021 (US). VEDVICK, Thomas, H.; 1301 Spring Street, Seattle, WA 98104 (US). TWARDZIK, Daniel, R.; 10195 South Beach Drive, Bainbridge Island, WA 98110 (US).		(54) Title: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY AND DIAGNOSIS OF TUBERCULOSIS																																									
<p>(57) Abstract</p> <p>Compounds and methods for inducing protective immunity against tuberculosis are disclosed. The compounds provided include polypeptides that contain at least one immunogenic portion of one or more <i>M. tuberculosis</i> proteins and DNA molecules encoding such polypeptides. Such compounds may be formulated into vaccines and/or pharmaceutical compositions for immunization against <i>M. tuberculosis</i> infection, or may be used for the diagnosis of tuberculosis.</p>																																											
<p style="text-align: center;"><b>D160 T Cell Proliferation</b></p> <table border="1"> <caption>Data for D160 T Cell Proliferation</caption> <thead> <tr> <th>Compound</th> <th>CPM Incorporated</th> </tr> </thead> <tbody> <tr> <td>Control</td> <td>~10</td> </tr> <tr> <td>Compound 1</td> <td>~100</td> </tr> <tr> <td>Compound 2</td> <td>~20</td> </tr> <tr> <td>Compound 3</td> <td>~15</td> </tr> </tbody> </table> <p style="text-align: center;"><b>D160 IFNg</b></p> <table border="1"> <caption>Data for D160 IFNg</caption> <thead> <tr> <th>Compound</th> <th>O.D. 450-570</th> </tr> </thead> <tbody> <tr> <td>Control</td> <td>~0.5</td> </tr> <tr> <td>Compound 1</td> <td>~5.5</td> </tr> <tr> <td>Compound 2</td> <td>~5.5</td> </tr> <tr> <td>Compound 3</td> <td>~2.5</td> </tr> </tbody> </table> <p style="text-align: center;"><b>D7 T Cell Proliferation</b></p> <table border="1"> <caption>Data for D7 T Cell Proliferation</caption> <thead> <tr> <th>Compound</th> <th>CPM Incorporated</th> </tr> </thead> <tbody> <tr> <td>Control</td> <td>~10</td> </tr> <tr> <td>Compound 1</td> <td>~100</td> </tr> <tr> <td>Compound 2</td> <td>~20</td> </tr> <tr> <td>Compound 3</td> <td>~15</td> </tr> </tbody> </table> <p style="text-align: center;"><b>D7 IFNg</b></p> <table border="1"> <caption>Data for D7 IFNg</caption> <thead> <tr> <th>Compound</th> <th>O.D. 450-570</th> </tr> </thead> <tbody> <tr> <td>Control</td> <td>~0.5</td> </tr> <tr> <td>Compound 1</td> <td>~5.5</td> </tr> <tr> <td>Compound 2</td> <td>~4.5</td> </tr> <tr> <td>Compound 3</td> <td>~2.5</td> </tr> </tbody> </table>				Compound	CPM Incorporated	Control	~10	Compound 1	~100	Compound 2	~20	Compound 3	~15	Compound	O.D. 450-570	Control	~0.5	Compound 1	~5.5	Compound 2	~5.5	Compound 3	~2.5	Compound	CPM Incorporated	Control	~10	Compound 1	~100	Compound 2	~20	Compound 3	~15	Compound	O.D. 450-570	Control	~0.5	Compound 1	~5.5	Compound 2	~4.5	Compound 3	~2.5
Compound	CPM Incorporated																																										
Control	~10																																										
Compound 1	~100																																										
Compound 2	~20																																										
Compound 3	~15																																										
Compound	O.D. 450-570																																										
Control	~0.5																																										
Compound 1	~5.5																																										
Compound 2	~5.5																																										
Compound 3	~2.5																																										
Compound	CPM Incorporated																																										
Control	~10																																										
Compound 1	~100																																										
Compound 2	~20																																										
Compound 3	~15																																										
Compound	O.D. 450-570																																										
Control	~0.5																																										
Compound 1	~5.5																																										
Compound 2	~4.5																																										
Compound 3	~2.5																																										

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LJ	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

Description5           COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
AND DIAGNOSIS OF TUBERCULOSISTechnical Field

10           The present invention relates generally to detecting, treating and preventing *Mycobacterium tuberculosis* infection. The invention is more particularly related to polypeptides comprising a *Mycobacterium tuberculosis* antigen, or a portion or other variant thereof, and the use of such polypeptides for diagnosing and vaccinating against *Mycobacterium tuberculosis* infection.

15

Background of the Invention

Tuberculosis is a chronic, infectious disease, that is generally caused by infection with *Mycobacterium tuberculosis*. It is a major disease in developing countries, as well as an increasing problem in developed areas of the world, with about 20 8 million new cases and 3 million deaths each year. Although the infection may be asymptomatic for a considerable period of time, the disease is most commonly manifested as an acute inflammation of the lungs, resulting in fever and a nonproductive cough. If left untreated, serious complications and death typically result.

Although tuberculosis can generally be controlled using extended 25 antibiotic therapy, such treatment is not sufficient to prevent the spread of the disease. Infected individuals may be asymptomatic, but contagious, for some time. In addition,

although compliance with the treatment regimen is critical, patient behavior is difficult to monitor. Some patients do not complete the course of treatment, which can lead to ineffective treatment and the development of drug resistance.

Inhibiting the spread of tuberculosis requires effective vaccination and  
5 accurate, early diagnosis of the disease. Currently, vaccination with live bacteria is the most efficient method for inducing protective immunity. The most common Mycobacterium employed for this purpose is *Bacillus Calmette-Guerin* (BCG), an avirulent strain of *Mycobacterium bovis*. However, the safety and efficacy of BCG is a source of controversy and some countries, such as the United States, do not vaccinate  
10 the general public. Diagnosis is commonly achieved using a skin test, which involves intradermal exposure to tuberculin PPD (protein-purified derivative). Antigen-specific T cell responses result in measurable induration at the injection site by 48-72 hours after injection, which indicates exposure to Mycobacterial antigens. Sensitivity and specificity have, however, been a problem with this test, and individuals vaccinated  
15 with BCG cannot be distinguished from infected individuals.

While macrophages have been shown to act as the principal effectors of *M. tuberculosis* immunity, T cells are the predominant inducers of such immunity. The essential role of T cells in protection against *M. tuberculosis* infection is illustrated by the frequent occurrence of *M. tuberculosis* in AIDS patients, due to the depletion of  
20 CD4 T cells associated with human immunodeficiency virus (HIV) infection. Mycobacterium-reactive CD4 T cells have been shown to be potent producers of gamma-interferon (IFN- $\gamma$ ), which, in turn, has been shown to trigger the anti-mycobacterial effects of macrophages in mice. While the role of IFN- $\gamma$  in humans is less clear, studies have shown that 1,25-dihydroxy-vitamin D3, either alone or in  
25 combination with IFN- $\gamma$  or tumor necrosis factor-alpha, activates human macrophages to inhibit *M. tuberculosis* infection. Furthermore, it is known that IFN- $\gamma$  stimulates human macrophages to make 1,25-dihydroxy-vitamin D3. Similarly, IL-12 has been shown to play a role in stimulating resistance to *M. tuberculosis* infection. For a review of the immunology of *M. tuberculosis* infection see Chan and Kaufmann in

*Tuberculosis: Pathogenesis, Protection and Control*, Bloom (ed.), ASM Press, Washington, DC, 1994.

Accordingly, there is a need in the art for improved vaccines and methods for preventing, treating and detecting tuberculosis. The present invention 5 fulfills these needs and further provides other related advantages.

#### Summary of the Invention

Briefly stated, this invention provides compounds and methods for preventing and diagnosing tuberculosis. In one aspect, polypeptides are provided 10 comprising an immunogenic portion of a soluble *M. tuberculosis* antigen, or a variant of such an antigen that differs only in conservative substitutions and/or modifications. In one embodiment of this aspect, the soluble antigen has one of the following N-terminal sequences:

- (a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Thr-Cys-Asn-Tyr-Gly-  
15 Gln-Val-Val-Ala-Ala-Leu; (SEQ ID No. 120)
- (b) Ala-Val-Glu-Ser-Gly-Met-Leu-Ala-Leu-Gly-Thr-Pro-Ala-Pro-  
Ser; (SEQ ID No. 121)
- (c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-  
Ala-Lys-Glu-Gly-Arg; (SEQ ID No. 122)
- 20 (d) Tyr-Tyr-Trp-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Trp-Gly-  
Pro; (SEQ ID No. 123)
- (e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val;  
(SEQ ID No. 124)
- 25 (f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro; (SEQ ID  
No. 125)
- (g) Asp-Pro-Glu-Pro-Ala-Pro-Pro-Val-Pro-Thr-Thr-Ala-Ala-Ser-  
Pro-Pro-Ser; (SEQ ID No. 126)
- (h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-  
Gly; (SEQ ID No. 127)

- (i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Leu-Thr-Ser-Leu-Leu-Asn-Ser-Leu-Ala-Asp-Pro-Asn-Val-Ser-Phe-Ala-Asn; (SEQ ID No. 128)
- 5 (j) Xaa-Asp-Ser-Glu-Lys-Ser-Ala-Thr-Ile-Lys-Val-Thr-Asp-Ala-Ser; (SEQ ID No. 134)
- (k) Ala-Gly-Asp-Thr-Xaa-Ile-Tyr-Ile-Val-Gly-Asn-Leu-Thr-Ala-Asp; (SEQ ID No. 135) or
- (l) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Leu-Gly-Gly-Thr-Val-Gln-Ala-Gly; (SEQ ID No. 136)

10 wherein Xaa may be any amino acid.

In a related aspect, polypeptides are provided comprising an immunogenic portion of an *M. tuberculosis* antigen, or a variant of such an antigen that differs only in conservative substitutions and/or modifications, the antigen having one of the following N-terminal sequences:

- 15 (m) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn-Val-His-Leu-Val; (SEQ ID No. 137) or
- (n) Asp-Pro-Pro-Asp-Pro-His-Gln-Xaa-Asp-Met-Thr-Lys-Gly-Tyr-Tyr-Pro-Gly-Gly-Arg-Arg-Xaa-Phe; (SEQ ID No. 129)

wherein Xaa may be any amino acid.

20 In another embodiment, the antigen comprises an amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID Nos.: 1, 2, 4-10, 13-25, 52, 99 and 101, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID Nos.: 1, 2, 4-10, 13-25, 52, 99 and 101 or a complement thereof under moderately stringent  
25 conditions.

In a related aspect, the polypeptides comprise an immunogenic portion of a *M. tuberculosis* antigen, or a variant of such an antigen that differs only in conservative substitutions and/or modifications, wherein the antigen comprises an amino acid sequence encoded by a DNA sequence selected from the group consisting of  
30 the sequences recited in SEQ ID Nos.: 26-51, the complements of said sequences, and

DNA sequences that hybridize to a sequence recited in SEQ ID Nos.: 26-51 or a complement thereof under moderately stringent conditions.

In related aspects, DNA sequences encoding the above polypeptides, expression vectors comprising these DNA sequences and host cells transformed or 5 transfected with such expression vectors are also provided.

In another aspect, the present invention provides fusion proteins comprising a first and a second inventive polypeptide or, alternatively, an inventive polypeptide and a known *M. tuberculosis* antigen.

Within other aspects, the present invention provides pharmaceutical 10 compositions that comprise one or more of the above polypeptides, or a DNA molecule encoding such polypeptides, and a physiologically acceptable carrier. The invention also provides vaccines comprising one or more of the polypeptides as described above and a non-specific immune response enhancer, together with vaccines comprising one or more DNA sequences encoding such polypeptides and a non-specific immune 15 response enhancer.

In yet another aspect, methods are provided for inducing protective immunity in a patient, comprising administering to a patient an effective amount of one or more of the above polypeptides.

In further aspects of this invention, methods and diagnostic kits are 20 provided for detecting tuberculosis in a patient. The methods comprise contacting dermal cells of a patient with one or more of the above polypeptides and detecting an immune response on the patient's skin. The diagnostic kits comprise one or more of the above polypeptides in combination with an apparatus sufficient to contact the polypeptide with the dermal cells of a patient.

25 These and other aspects of the present invention will become apparent upon reference to the following detailed description and attached drawings. All references disclosed herein are hereby incorporated by reference in their entirety as if each was incorporated individually.

Brief Description of the Drawings and Sequence Identifiers

Figure 1A and B illustrate the stimulation of proliferation and interferon- $\gamma$  production in T cells derived from a first and a second *M. tuberculosis*-immune donor, respectively, by the 14 Kd, 20 Kd and 26 Kd antigens described in Example 1.

5       Figure 2 illustrates the stimulation of proliferation and interferon- $\gamma$  production in T cells derived from an *M. tuberculosis*-immune individual by the two representative polypeptides TbRa3 and TbRa9.

- SEQ. ID NO. 1 is the DNA sequence of TbRa1.  
10      SEQ. ID NO. 2 is the DNA sequence of TbRa10.  
SEQ. ID NO. 3 is the DNA sequence of TbRa11.  
SEQ. ID NO. 4 is the DNA sequence of TbRa12.  
SEQ. ID NO. 5 is the DNA sequence of TbRa13.  
SEQ. ID NO. 6 is the DNA sequence of TbRa16.  
15      SEQ. ID NO. 7 is the DNA sequence of TbRa17.  
SEQ. ID NO. 8 is the DNA sequence of TbRa18.  
SEQ. ID NO. 9 is the DNA sequence of TbRa19.  
SEQ. ID NO. 10 is the DNA sequence of TbRa24.  
SEQ. ID NO. 11 is the DNA sequence of TbRa26.  
20      SEQ. ID NO. 12 is the DNA sequence of TbRa28.  
SEQ. ID NO. 13 is the DNA sequence of TbRa29.  
SEQ. ID NO. 14 is the DNA sequence of TbRa2A.  
SEQ. ID NO. 15 is the DNA sequence of TbRa3.  
SEQ. ID NO. 16 is the DNA sequence of TbRa32.  
25      SEQ. ID NO. 17 is the DNA sequence of TbRa35.  
SEQ. ID NO. 18 is the DNA sequence of TbRa36.  
SEQ. ID NO. 19 is the DNA sequence of TbRa4.  
SEQ. ID NO. 20 is the DNA sequence of TbRa9.  
SEQ. ID NO. 21 is the DNA sequence of TbRaB.  
30      SEQ. ID NO. 22 is the DNA sequence of TbRaC.

- SEQ. ID NO. 23 is the DNA sequence of TbRaD.
- SEQ. ID NO. 24 is the DNA sequence of YYWCPG.
- SEQ. ID NO. 25 is the DNA sequence of AAMK.
- SEQ. ID NO. 26 is the DNA sequence of TbL-23.
- 5 SEQ. ID NO. 27 is the DNA sequence of TbL-24.
- SEQ. ID NO. 28 is the DNA sequence of TbL-25.
- SEQ. ID NO. 29 is the DNA sequence of TbL-28.
- SEQ. ID NO. 30 is the DNA sequence of TbL-29.
- SEQ. ID NO. 31 is the DNA sequence of TbH-5.
- 10 SEQ. ID NO. 32 is the DNA sequence of TbH-8.
- SEQ. ID NO. 33 is the DNA sequence of TbH-9.
- SEQ. ID NO. 34 is the DNA sequence of TbM-1.
- SEQ. ID NO. 35 is the DNA sequence of TbM-3.
- SEQ. ID NO. 36 is the DNA sequence of TbM-6.
- 15 SEQ. ID NO. 37 is the DNA sequence of TbM-7.
- SEQ. ID NO. 38 is the DNA sequence of TbM-9.
- SEQ. ID NO. 39 is the DNA sequence of TbM-12.
- SEQ. ID NO. 40 is the DNA sequence of TbM-13.
- SEQ. ID NO. 41 is the DNA sequence of TbM-14.
- 20 SEQ. ID NO. 42 is the DNA sequence of TbM-15.
- SEQ. ID NO. 43 is the DNA sequence of TbH-4.
- SEQ. ID NO. 44 is the DNA sequence of TbH-4-FWD.
- SEQ. ID NO. 45 is the DNA sequence of TbH-12.
- SEQ. ID NO. 46 is the DNA sequence of Tb38-1.
- 25 SEQ. ID NO. 47 is the DNA sequence of Tb38-4.
- SEQ. ID NO. 48 is the DNA sequence of TbL-17.
- SEQ. ID NO. 49 is the DNA sequence of TbL-20.
- SEQ. ID NO. 50 is the DNA sequence of TbL-21.
- SEQ. ID NO. 51 is the DNA sequence of TbH-16.
- 30 SEQ. ID NO. 52 is the DNA sequence of DPEP.

- SEQ. ID NO. 53 is the deduced amino acid sequence of DPEP.
- SEQ. ID NO. 54 is the protein sequence of DPV N-terminal Antigen.
- SEQ. ID NO. 55 is the protein sequence of AVGS N-terminal Antigen.
- SEQ. ID NO. 56 is the protein sequence of AAMK N-terminal Antigen.
- 5 SEQ. ID NO. 57 is the protein sequence of YYWC N-terminal Antigen.
- SEQ. ID NO. 58 is the protein sequence of DIGS N-terminal Antigen.
- SEQ. ID NO. 59 is the protein sequence of AEES N-terminal Antigen.
- SEQ. ID NO. 60 is the protein sequence of DPEP N-terminal Antigen.
- SEQ. ID NO. 61 is the protein sequence of APKT N-terminal Antigen.
- 10 SEQ. ID NO. 62 is the protein sequence of DPAS N-terminal Antigen.
- SEQ. ID NO. 63 is the deduced amino acid sequence of TbRa1.
- SEQ. ID NO. 64 is the deduced amino acid sequence of TbRa10.
- SEQ. ID NO. 65 is the deduced amino acid sequence of TbRa11.
- SEQ. ID NO. 66 is the deduced amino acid sequence of TbRa12.
- 15 SEQ. ID NO. 67 is the deduced amino acid sequence of TbRa13.
- SEQ. ID NO. 68 is the deduced amino acid sequence of TbRa16.
- SEQ. ID NO. 69 is the deduced amino acid sequence of TbRa17.
- SEQ. ID NO. 70 is the deduced amino acid sequence of TbRa18.
- SEQ. ID NO. 71 is the deduced amino acid sequence of TbRa19.
- 20 SEQ. ID NO. 72 is the deduced amino acid sequence of TbRa24.
- SEQ. ID NO. 73 is the deduced amino acid sequence of TbRa26.
- SEQ. ID NO. 74 is the deduced amino acid sequence of TbRa28.
- SEQ. ID NO. 75 is the deduced amino acid sequence of TbRa29.
- SEQ. ID NO. 76 is the deduced amino acid sequence of TbRa2A.
- 25 SEQ. ID NO. 77 is the deduced amino acid sequence of TbRa3.
- SEQ. ID NO. 78 is the deduced amino acid sequence of TbRa32.
- SEQ. ID NO. 79 is the deduced amino acid sequence of TbRa35.
- SEQ. ID NO. 80 is the deduced amino acid sequence of TbRa36.
- SEQ. ID NO. 81 is the deduced amino acid sequence of TbRa4.
- 30 SEQ. ID NO. 82 is the deduced amino acid sequence of TbRa9.

- SEQ. ID NO. 83 is the deduced amino acid sequence of TbRaB.  
SEQ. ID NO. 84 is the deduced amino acid sequence of TbRaC.  
SEQ. ID NO. 85 is the deduced amino acid sequence of TbRaD.  
SEQ. ID NO. 86 is the deduced amino acid sequence of YYWCPG.  
5 SEQ. ID NO. 87 is the deduced amino acid sequence of TbAAMK.  
SEQ. ID NO. 88 is the deduced amino acid sequence of Tb38-1.  
SEQ. ID NO. 89 is the deduced amino acid sequence of TbH-4.  
SEQ. ID NO. 90 is the deduced amino acid sequence of TbH-8.  
SEQ. ID NO. 91 is the deduced amino acid sequence of TbH-9.  
10 SEQ. ID NO. 92 is the deduced amino acid sequence of TbH-12.  
SEQ. ID NO. 93 is the amino acid sequence of Tb38-1 Peptide 1.  
SEQ. ID NO. 94 is the amino acid sequence of Tb38-1 Peptide 2.  
SEQ. ID NO. 95 is the amino acid sequence of Tb38-1 Peptide 3.  
SEQ. ID NO. 96 is the amino acid sequence of Tb38-1 Peptide 4.  
15 SEQ. ID NO. 97 is the amino acid sequence of Tb38-1 Peptide 5.  
SEQ. ID NO. 98 is the amino acid sequence of Tb38-1 Peptide 6.  
SEQ. ID NO. 99 is the DNA sequence of DPAS.  
SEQ. ID NO. 100 is the deduced amino acid sequence of DPAS.  
SEQ. ID NO. 101 is the DNA sequence of DPV.  
20 SEQ. ID NO. 102 is the deduced amino acid sequence of DPV.  
SEQ. ID NO. 103 is the DNA sequence of ESAT-6.  
SEQ. ID NO. 104 is the deduced amino acid sequence of ESAT-6.  
SEQ. ID NO. 105 is the DNA sequence of TbH-8-2.  
SEQ. ID NO. 106 is the DNA sequence of TbH-9FL.  
25 SEQ. ID NO. 107 is the deduced amino acid sequence of TbH-9FL.  
SEQ. ID NO. 108 is the DNA sequence of TbH-9-1.  
SEQ. ID NO. 109 is the deduced amino acid sequence of TbH-9-1.  
SEQ. ID NO. 110 is the DNA sequence of TbH-9-4.  
SEQ. ID NO. 111 is the deduced amino acid sequence of TbH-9-4.  
30 SEQ. ID NO. 112 is the DNA sequence of Tb38-1F2 IN.

- SEQ. ID NO. 113 is the DNA sequence of Tb38-2F2 RP.
- SEQ. ID NO. 114 is the deduced amino acid sequence of Tb37-FL.
- SEQ. ID NO. 115 is the deduced amino acid sequence of Tb38-IN.
- SEQ. ID NO. 116 is the DNA sequence of Tb38-1F3.
- 5 SEQ. ID NO. 117 is the deduced amino acid sequence of Tb38-1F3.
- SEQ. ID NO. 118 is the DNA sequence of Tb38-1F5.
- SEQ. ID NO. 119 is the DNA sequence of Tb38-1F6.
- SEQ. ID NO. 120 is the deduced N-terminal amino acid sequence of DPV.
- SEQ. ID NO. 121 is the deduced N-terminal amino acid sequence of AVGS.
- 10 SEQ. ID NO. 122 is the deduced N-terminal amino acid sequence of AAMK.
- SEQ. ID NO. 123 is the deduced N-terminal amino acid sequence of YYWC.
- SEQ. ID NO. 124 is the deduced N-terminal amino acid sequence of DIGS.
- SEQ. ID NO. 125 is the deduced N-terminal amino acid sequence of AEES.
- SEQ. ID NO. 126 is the deduced N-terminal amino acid sequence of DPEP.
- 15 SEQ. ID NO. 127 is the deduced N-terminal amino acid sequence of APK7.
- SEQ. ID NO. 128 is the deduced amino acid sequence of DPAS.
- SEQ. ID NO. 129 is the protein sequence of DPPD N-terminal Antigen.
- SEQ ID NO. 130-133 are the protein sequences of four DPPD cyanogen bromide fragments.
- 20 SEQ ID NO. 134 is the N-terminal protein sequence of XDS antigen.
- SEQ ID NO. 135 is the N-terminal protein sequence of AGD antigen.
- SEQ ID NO. 136 is the N-terminal protein sequence of APE antigen.
- SEQ ID NO. 137 is the N-terminal protein sequence of XYI antigen.

25 Detailed Description of the Invention

As noted above, the present invention is generally directed to compositions and methods for preventing, treating and diagnosing tuberculosis. The compositions of the subject invention include polypeptides that comprise at least one immunogenic portion of a *M. tuberculosis* antigen, or a variant of such an antigen that 30 differs only in conservative substitutions and/or modifications. Polypeptides within the scope of the present invention include, but are not limited to, immunogenic soluble

*M. tuberculosis* antigens. A "soluble *M. tuberculosis* antigen" is a protein of *M. tuberculosis* origin that is present in *M. tuberculosis* culture filtrate. As used herein, the term "polypeptide" encompasses amino acid chains of any length, including full length proteins (*i.e.*, antigens), wherein the amino acid residues are linked by covalent peptide bonds. Thus, a polypeptide comprising an immunogenic portion of one of the above antigens may consist entirely of the immunogenic portion, or may contain additional sequences. The additional sequences may be derived from the native *M. tuberculosis* antigen or may be heterologous, and such sequences may (but need not) be immunogenic.

10        "Immunogenic," as used herein, refers to the ability to elicit an immune response (*e.g.*, cellular) in a patient, such as a human, and/or in a biological sample. In particular, antigens that are immunogenic (and immunogenic portions or other variants of such antigens) are capable of stimulating cell proliferation, interleukin-12 production and/or interferon- $\gamma$  production in biological samples comprising one or more cells selected from the group of T cells, NK cells, B cells and macrophages, where the cells are derived from an *M. tuberculosis*-immune individual. Polypeptides comprising at least an immunogenic portion of one or more *M. tuberculosis* antigens may generally be used to detect tuberculosis or to induce protective immunity against tuberculosis in a patient.

15        The compositions and methods of this invention also encompass variants of the above polypeptides. A "variant," as used herein, is a polypeptide that differs from the native antigen only in conservative substitutions and/or modifications, such that the ability of the polypeptide to induce an immune response is retained. Such variants may generally be identified by modifying one of the above polypeptide sequences, and evaluating the immunogenic properties of the modified polypeptide using, for example, the representative procedures described herein.

20        A "conservative substitution" is one in which an amino acid is substituted for another amino acid that has similar properties, such that one skilled in the art of peptide chemistry would expect the secondary structure and hydrophobic nature of the polypeptide to be substantially unchanged. In general, the following

groups of amino acids represent conservative changes: (1) ala, pro, gly, glu, asp, gln, asn, ser, thr; (2) cys, ser, tyr, thr; (3) val, ile, leu, met, ala, phe; (4) lys, arg, his; and (5) phe, tyr, trp, his.

Variants may also (or alternatively) be modified by, for example, the  
5 deletion or addition of amino acids that have minimal influence on the immunogenic properties, secondary structure and hydrophobic nature of the polypeptide. For example, a polypeptide may be conjugated to a signal (or leader) sequence at the N-terminal end of the protein which co-translationally or post-translationally directs transfer of the protein. The polypeptide may also be conjugated to a linker or other sequence for ease  
10 of synthesis, purification or identification of the polypeptide (e.g., poly-His), or to enhance binding of the polypeptide to a solid support. For example, a polypeptide may be conjugated to an immunoglobulin Fc region.

In a related aspect, combination polypeptides are disclosed. A "combination polypeptide" is a polypeptide comprising at least one of the above  
15 immunogenic portions and one or more additional immunogenic *M. tuberculosis* sequences, which are joined via a peptide linkage into a single amino acid chain. The sequences may be joined directly (*i.e.*, with no intervening amino acids) or may be joined by way of a linker sequence (*e.g.*, Gly-Cys-Gly) that does not significantly diminish the immunogenic properties of the component polypeptides.

20 In general, *M. tuberculosis* antigens, and DNA sequences encoding such antigens, may be prepared using any of a variety of procedures. For example, soluble antigens may be isolated from *M. tuberculosis* culture filtrate by procedures known to those of ordinary skill in the art, including anion-exchange and reverse phase chromatography. Purified antigens are then evaluated for their ability to elicit an  
25 appropriate immune response (*e.g.*, cellular) using, for example, the representative methods described herein. Immunogenic antigens may then be partially sequenced using techniques such as traditional Edman chemistry. *See* Edman and Berg, *Eur. J. Biochem.* 80:116-132, 1967.

Immunogenic antigens may also be produced recombinantly using a  
30 DNA sequence that encodes the antigen, which has been inserted into an expression

- vector and expressed in an appropriate host. DNA molecules encoding soluble antigens may be isolated by screening an appropriate *M. tuberculosis* expression library with anti-sera (*e.g.*, rabbit) raised specifically against soluble *M. tuberculosis* antigens. DNA sequences encoding antigens that may or may not be soluble may be identified by
- 5 screening an appropriate *M. tuberculosis* genomic or cDNA expression library with sera obtained from patients infected with *M. tuberculosis*. Such screens may generally be performed using techniques well known to those of ordinary skill in the art, such as those described in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989.
- 10 DNA sequences encoding soluble antigens may also be obtained by screening an appropriate *M. tuberculosis* cDNA or genomic DNA library for DNA sequences that hybridize to degenerate oligonucleotides derived from partial amino acid sequences of isolated soluble antigens. Degenerate oligonucleotide sequences for use in such a screen may be designed and synthesized, and the screen may be performed, as
- 15 described (for example) in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989 (and references cited therein). Polymerase chain reaction (PCR) may also be employed, using the above oligonucleotides in methods well known in the art, to isolate a nucleic acid probe from a cDNA or genomic library. The library screen may then be performed using the isolated
- 20 probe.

Alternatively, genomic or cDNA libraries derived from *M. tuberculosis* may be screened directly using peripheral blood mononuclear cells (PBMCs) or T cell lines or clones derived from one or more *M. tuberculosis*-immune individuals. In general, PBMCs and/or T cells for use in such screens may be prepared as described

25 below. Direct library screens may generally be performed by assaying pools of expressed recombinant proteins for the ability to induce proliferation and/or interferon- $\gamma$  production in T cells derived from an *M. tuberculosis*-immune individual. Alternatively, potential T cell antigens may be first selected based on antibody reactivity, as described above.

Regardless of the method of preparation, the antigens (and immunogenic portions thereof) described herein (which may or may not be soluble) have the ability to induce an immunogenic response. More specifically, the antigens have the ability to induce proliferation and/or cytokine production (*i.e.*, interferon- $\gamma$  and/or interleukin-12 production) in T cells, NK cells, B cells and/or macrophages derived from an *M. tuberculosis*-immune individual. The selection of cell type for use in evaluating an immunogenic response to a antigen will, of course, depend on the desired response. For example, interleukin-12 production is most readily evaluated using preparations containing B cells and/or macrophages. An *M. tuberculosis*-immune individual is one who is considered to be resistant to the development of tuberculosis by virtue of having mounted an effective T cell response to *M. tuberculosis* (*i.e.*, substantially free of disease symptoms). Such individuals may be identified based on a strongly positive (*i.e.*, greater than about 10 mm diameter induration) intradermal skin test response to tuberculosis proteins (PPD) and an absence of any signs or symptoms of tuberculosis disease. T cells, NK cells, B cells and macrophages derived from *M. tuberculosis*-immune individuals may be prepared using methods known to those of ordinary skill in the art. For example, a preparation of PBMCs (*i.e.*, peripheral blood mononuclear cells) may be employed without further separation of component cells. PBMCs may generally be prepared, for example, using density centrifugation through Ficoll™ (Winthrop Laboratories, NY). T cells for use in the assays described herein may also be purified directly from PBMCs. Alternatively, an enriched T cell line reactive against mycobacterial proteins, or T cell clones reactive to individual mycobacterial proteins, may be employed. Such T cell clones may be generated by, for example, culturing PBMCs from *M. tuberculosis*-immune individuals with mycobacterial proteins for a period of 2-4 weeks. This allows expansion of only the mycobacterial protein-specific T cells, resulting in a line composed solely of such cells. These cells may then be cloned and tested with individual proteins, using methods known to those of ordinary skill in the art, to more accurately define individual T cell specificity. In general, antigens that test positive in assays for proliferation and/or cytokine production (*i.e.*, interferon- $\gamma$  and/or interleukin-12 production) performed using T cells, NK cells, B cells

and/or macrophages derived from an *M. tuberculosis*-immune individual are considered immunogenic. Such assays may be performed, for example, using the representative procedures described below. Immunogenic portions of such antigens may be identified using similar assays, and may be present within the polypeptides described herein.

- 5       The ability of a polypeptide (e.g., an immunogenic antigen, or a portion or other variant thereof) to induce cell proliferation is evaluated by contacting the cells (e.g., T cells and/or NK cells) with the polypeptide and measuring the proliferation of the cells. In general, the amount of polypeptide that is sufficient for evaluation of about  $10^5$  cells ranges from about 10 ng/mL to about 100  $\mu$ g/mL and preferably is about  
10      10  $\mu$ g/mL. The incubation of polypeptide with cells is typically performed at 37°C for about six days. Following incubation with polypeptide, the cells are assayed for a proliferative response, which may be evaluated by methods known to those of ordinary skill in the art, such as exposing cells to a pulse of radiolabeled thymidine and measuring the incorporation of label into cellular DNA. In general, a polypeptide that  
15     results in at least a three fold increase in proliferation above background (*i.e.*, the proliferation observed for cells cultured without polypeptide) is considered to be able to induce proliferation.

- 20       The ability of a polypeptide to stimulate the production of interferon- $\gamma$  and/or interleukin-12 in cells may be evaluated by contacting the cells with the polypeptide and measuring the level of interferon- $\gamma$  or interleukin-12 produced by the cells. In general, the amount of polypeptide that is sufficient for the evaluation of about  $10^5$  cells ranges from about 10 ng/mL to about 100  $\mu$ g/mL and preferably is about 10  $\mu$ g/mL. The polypeptide may, but need not, be immobilized on a solid support, such as a bead or a biodegradable microsphere, such as those described in U.S. Patent  
25     Nos. 4,897,268 and 5,075,109. The incubation of polypeptide with the cells is typically performed at 37°C for about six days. Following incubation with polypeptide, the cells are assayed for interferon- $\gamma$  and/or interleukin-12 (or one or more subunits thereof), which may be evaluated by methods known to those of ordinary skill in the art, such as an enzyme-linked immunosorbent assay (ELISA) or, in the case of IL-12 P70 subunit, a  
30     bioassay such as an assay measuring proliferation of T cells. In general, a polypeptide

that results in the production of at least 50 pg of interferon- $\gamma$  per mL of cultured supernatant (containing  $10^4$ - $10^5$  T cells per mL) is considered able to stimulate the production of interferon- $\gamma$ . A polypeptide that stimulates the production of at least 10 pg/mL of IL-12 P70 subunit, and/or at least 100 pg/mL of IL-12 P40 subunit, per  $10^5$  5 macrophages or B cells (or per  $3 \times 10^5$  PBMC) is considered able to stimulate the production of IL-12.

In general, immunogenic antigens are those antigens that stimulate proliferation and/or cytokine production (*i.e.*, interferon- $\gamma$  and/or interleukin-12 production) in T cells, NK cells, B cells and/or macrophages derived from at least about 10 25% of *M. tuberculosis*-immune individuals. Among these immunogenic antigens, polypeptides having superior therapeutic properties may be distinguished based on the magnitude of the responses in the above assays and based on the percentage of individuals for which a response is observed. In addition, antigens having superior therapeutic properties will not stimulate proliferation and/or cytokine production *in* 15 *vitro* in cells derived from more than about 25% of individuals that are not *M. tuberculosis*-immune, thereby eliminating responses that are not specifically due to *M. tuberculosis*-responsive cells. Those antigens that induce a response in a high percentage of T cell, NK cell, B cell and/or macrophage preparations from 20 *M. tuberculosis*-immune individuals (with a low incidence of responses in cell preparations from other individuals) have superior therapeutic properties.

Antigens with superior therapeutic properties may also be identified based on their ability to diminish the severity of *M. tuberculosis* infection in experimental animals, when administered as a vaccine. Suitable vaccine preparations for use on experimental animals are described in detail below. Efficacy may be 25 determined based on the ability of the antigen to provide at least about a 50% reduction in bacterial numbers and/or at least about a 40% decrease in mortality following experimental infection. Suitable experimental animals include mice, guinea pigs and primates.

Antigens having superior diagnostic properties may generally be 30 identified based on the ability to elicit a response in an intradermal skin test performed

on an individual with active tuberculosis, but not in a test performed on an individual who is not infected with *M. tuberculosis*. Skin tests may generally be performed as described below, with a response of at least 5 mm induration considered positive.

Immunogenic portions of the antigens described herein may be prepared and identified using well known techniques, such as those summarized in Paul, *Fundamental Immunology*, 3d ed., Raven Press, 1993, pp. 243-247 and references cited therein. Such techniques include screening polypeptide portions of the native antigen for immunogenic properties. The representative proliferation and cytokine production assays described herein may generally be employed in these screens. An immunogenic portion of a polypeptide is a portion that, within such representative assays, generates an immune response (e.g., proliferation, interferon- $\gamma$  production and/or interleukin-12 production) that is substantially similar to that generated by the full length antigen. In other words, an immunogenic portion of an antigen may generate at least about 20%, and preferably about 100%, of the proliferation induced by the full length antigen in the model proliferation assay described herein. An immunogenic portion may also, or alternatively, stimulate the production of at least about 20%, and preferably about 100%, of the interferon- $\gamma$  and/or interleukin-12 induced by the full length antigen in the model assay described herein.

Portions and other variants of *M. tuberculosis* antigens may be generated by synthetic or recombinant means. Synthetic polypeptides having fewer than about 100 amino acids, and generally fewer than about 50 amino acids, may be generated using techniques well known to those of ordinary skill in the art. For example, such polypeptides may be synthesized using any of the commercially available solid-phase techniques, such as the Merrifield solid-phase synthesis method, where amino acids are sequentially added to a growing amino acid chain. See Merrifield, *J. Am. Chem. Soc.* 85:2149-2146, 1963. Equipment for automated synthesis of polypeptides is commercially available from suppliers such as Applied BioSystems, Inc., Foster City, CA, and may be operated according to the manufacturer's instructions. Variants of a native antigen may generally be prepared using standard mutagenesis techniques, such as oligonucleotide-directed site-specific mutagenesis. Sections of the DNA sequence

may also be removed using standard techniques to permit preparation of truncated polypeptides.

Recombinant polypeptides containing portions and/or variants of a native antigen may be readily prepared from a DNA sequence encoding the polypeptide 5 using a variety of techniques well known to those of ordinary skill in the art. For example, supernatants from suitable host/vector systems which secrete recombinant protein into culture media may be first concentrated using a commercially available filter. Following concentration, the concentrate may be applied to a suitable purification matrix such as an affinity matrix or an ion exchange resin. Finally, one or 10 more reverse phase HPLC steps can be employed to further purify a recombinant protein.

Any of a variety of expression vectors known to those of ordinary skill in the art may be employed to express recombinant polypeptides of this invention. Expression may be achieved in any appropriate host cell that has been transformed or 15 transfected with an expression vector containing a DNA molecule that encodes a recombinant polypeptide. Suitable host cells include prokaryotes, yeast and higher eukaryotic cells. Preferably, the host cells employed are *E. coli*, yeast or a mammalian cell line such as COS or CHO. The DNA sequences expressed in this manner may encode naturally occurring antigens, portions of naturally occurring antigens, or other 20 variants thereof.

In general, regardless of the method of preparation, the polypeptides disclosed herein are prepared in substantially pure form. Preferably, the polypeptides are at least about 80% pure, more preferably at least about 90% pure and most preferably at least about 99% pure. In certain preferred embodiments, described in 25 detail below, the substantially pure polypeptides are incorporated into pharmaceutical compositions or vaccines for use in one or more of the methods disclosed herein.

In certain specific embodiments, the subject invention discloses polypeptides comprising at least an immunogenic portion of a soluble *M. tuberculosis* antigen having one of the following N-terminal sequences, or a variant thereof that 30 differs only in conservative substitutions and/or modifications:

- (a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Thr-Cys-Asn-Tyr-Gly-Gln-Val-Val-Ala-Ala-Leu; (SEQ ID No. 120)
- 5 (b) Ala-Val-Glu-Ser-Gly-Met-Leu-Ala-Leu-Gly-Thr-Pro-Ala-Pro-Ser; (SEQ ID No. 121)
- (c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-Ala-Lys-Glu-Gly-Arg; (SEQ ID No. 122)
- (d) Tyr-Tyr-Trp-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Trp-Gly-Pro; (SEQ ID No. 123)
- 10 (e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val; (SEQ ID No. 124)
- (f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro; (SEQ ID No. 125)
- (g) Asp-Pro-Glu-Pro-Ala-Pro-Pro-Val-Pro-Thr-Ala-Ala-Ala-Ser-Pro-Pro-Ser; (SEQ ID No. 126)
- 15 (h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly; (SEQ ID No. 127)
- (i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Leu-Thr-Ser-Leu-Leu-Asn-Ser-Leu-Ala-Asp-Pro-Asn-Val-Ser-Phe-Ala-Asn; (SEQ ID No. 128)
- 20 (j) Xaa-Asp-Ser-Glu-Lys-Ser-Ala-Thr-Ile-Lys-Val-Thr-Asp-Ala-Ser; (SEQ ID No. 134)
- (k) Ala-Gly-Asp-Thr-Xaa-Ile-Tyr-Ile-Val-Gly-Asn-Leu-Thr-Ala-Asp; (SEQ ID No. 135) or
- (l) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Leu-Gly-Gly-Thr-Val-Gln-Ala-Gly; (SEQ ID No. 136)

25 wherein Xaa may be any amino acid, preferably a cysteine residue. A DNA sequence encoding the antigen identified as (g) above is provided in SEQ ID No. 52, and the polypeptide encoded by SEQ ID No. 52 is provided in SEQ ID No. 53. A DNA sequence encoding the antigen defined as (a) above is provided in SEQ ID No. 101; its deduced amino acid sequence is provided in SEQ ID No. 102. A DNA sequence

corresponding to antigen (d) above is provided in SEQ ID No. 24 a DNA sequence corresponding to antigen (c) is provided in SEQ ID No. 25 and a DNA sequence corresponding to antigen (i) is provided in SEQ ID No. 99; its deduced amino acid sequence is provided in SEQ ID No. 100.

5 In a further specific embodiment, the subject invention discloses polypeptides comprising at least an immunogenic portion of an *M. tuberculosis* antigen having one of the following N-terminal sequences, or a variant thereof that differs only in conservative substitutions and/or modifications:

- 10 (m) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn-Val-His-Leu-Val; (SEQ ID No 137) or  
(n) Asp-Pro-Pro-Asp-Pro-His-Gln-Xaa-Asp-Met-Thr-Lys-Gly-Tyr-Tyr-Pro-Gly-Gly-Arg-Arg-Xaa-Phe; (SEQ ID No. 129)

wherein Xaa may be any amino acid, preferably a cysteine residue.

15 In other specific embodiments, the subject invention discloses polypeptides comprising at least an immunogenic portion of a soluble *M. tuberculosis* antigen (or a variant of such an antigen) that comprises one or more of the amino acid sequences encoded by (a) the DNA sequences of SEQ ID Nos.: 1, 2, 4-10, 13-25 and 52; (b) the complements of such DNA sequences, or (c) DNA sequences substantially homologous to a sequence in (a) or (b).

20 25 In further specific embodiments, the subject invention discloses polypeptides comprising at least an immunogenic portion of a *M. tuberculosis* antigen (or a variant of such an antigen), which may or may not be soluble, that comprises one or more of the amino acid sequences encoded by (a) the DNA sequences of SEQ ID Nos.: 26-51, (b) the complements of such DNA sequences or (c) DNA sequences substantially homologous to a sequence in (a) or (b).

In the specific embodiments discussed above, the *M. tuberculosis* antigens include variants that are encoded by DNA sequences which are substantially homologous to one or more of DNA sequences specifically recited herein. "Substantial homology," as used herein, refers to DNA sequences that are capable of hybridizing under moderately stringent conditions. Suitable moderately stringent conditions include

prewashing in a solution of 5X SSC, 0.5% SDS, 1.0 mM EDTA (pH 8.0); hybridizing at 50°C-65°C, 5X SSC, overnight or, in the case of cross-species homology at 45°C, 0.5X SSC; followed by washing twice at 65°C for 20 minutes with each of 2X, 0.5X and 0.2X SSC containing 0.1% SDS). Such hybridizing DNA sequences are also 5 within the scope of this invention, as are nucleotide sequences that, due to code degeneracy, encode an immunogenic polypeptide that is encoded by a hybridizing DNA sequence.

In a related aspect, the present invention provides fusion proteins comprising a first and a second inventive polypeptide or, alternatively, a polypeptide of 10 the present invention and a known *M. tuberculosis* antigen, such as the 38 kD antigen described above or ESAT-6 (SEQ ID Nos. 103 and 104), together with variants of such fusion proteins. The fusion proteins of the present invention may also include a linker peptide between the first and second polypeptides.

A DNA sequence encoding a fusion protein of the present invention is 15 constructed using known recombinant DNA techniques to assemble separate DNA sequences encoding the first and second polypeptides into an appropriate expression vector. The 3' end of a DNA sequence encoding the first polypeptide is ligated, with or without a peptide linker, to the 5' end of a DNA sequence encoding the second polypeptide so that the reading frames of the sequences are in phase to permit mRNA 20 translation of the two DNA sequences into a single fusion protein that retains the biological activity of both the first and the second polypeptides.

A peptide linker sequence may be employed to separate the first and the second polypeptides by a distance sufficient to ensure that each polypeptide folds into its secondary and tertiary structures. Such a peptide linker sequence is incorporated into 25 the fusion protein using standard techniques well known in the art. Suitable peptide linker sequences may be chosen based on the following factors: (1) their ability to adopt a flexible extended conformation; (2) their inability to adopt a secondary structure that could interact with functional epitopes on the first and second polypeptides; and (3) the lack of hydrophobic or charged residues that might react with the polypeptide 30 functional epitopes. Preferred peptide linker sequences contain Gly, Asn and Ser

residues. Other near neutral amino acids, such as Thr and Ala may also be used in the linker sequence. Amino acid sequences which may be usefully employed as linkers include those disclosed in Maratea et al., *Gene* 40:39-46, 1985; Murphy et al., *Proc. Natl. Acad. Sci. USA* 83:8258-8262, 1986; U.S. Patent No. 4,935,233 and U.S. Patent 5 No. 4,751,180. The linker sequence may be from 1 to about 50 amino acids in length. Peptide sequences are not required when the first and second polypeptides have non-essential N-terminal amino acid regions that can be used to separate the functional domains and prevent steric interference.

The ligated DNA sequences are operably linked to suitable 10 transcriptional or translational regulatory elements. The regulatory elements responsible for expression of DNA are located only 5' to the DNA sequence encoding the first polypeptides. Similarly, stop codons require to end translation and transcription termination signals are only present 3' to the DNA sequence encoding the second polypeptide.

15 In another aspect, the present invention provides methods for using one or more of the above polypeptides or fusion proteins (or DNA molecules encoding such polypeptides) to induce protective immunity against tuberculosis in a patient. As used herein, a "patient" refers to any warm-blooded animal, preferably a human. A patient may be afflicted with a disease, or may be free of detectable disease and/or infection. In 20 other words, protective immunity may be induced to prevent or treat tuberculosis.

In this aspect, the polypeptide, fusion protein or DNA molecule is generally present within a pharmaceutical composition and/or a vaccine. Pharmaceutical compositions may comprise one or more polypeptides, each of which may contain one or more of the above sequences (or variants thereof), and a 25 physiologically acceptable carrier. Vaccines may comprise one or more of the above polypeptides and a non-specific immune response enhancer, such as an adjuvant or a liposome (into which the polypeptide is incorporated). Such pharmaceutical compositions and vaccines may also contain other *M. tuberculosis* antigens, either incorporated into a combination polypeptide or present within a separate polypeptide.

Alternatively, a vaccine may contain DNA encoding one or more polypeptides as described above, such that the polypeptide is generated *in situ*. In such vaccines, the DNA may be present within any of a variety of delivery systems known to those of ordinary skill in the art, including nucleic acid expression systems, bacterial and viral expression systems. Appropriate nucleic acid expression systems contain the necessary DNA sequences for expression in the patient (such as a suitable promoter and terminating signal). Bacterial delivery systems involve the administration of a bacterium (such as *Bacillus-Calmette-Guerrin*) that expresses an immunogenic portion of the polypeptide on its cell surface. In a preferred embodiment, the DNA may be introduced using a viral expression system (e.g., vaccinia or other pox virus, retrovirus, or adenovirus), which may involve the use of a non-pathogenic (defective), replication competent virus. Techniques for incorporating DNA into such expression systems are well known to those of ordinary skill in the art. The DNA may also be "naked," as described, for example, in Ulmer et al., *Science* 259:1745-1749, 1993 and reviewed by Cohen, *Science* 259:1691-1692, 1993. The uptake of naked DNA may be increased by coating the DNA onto biodegradable beads, which are efficiently transported into the cells.

In a related aspect, a DNA vaccine as described above may be administered simultaneously with or sequentially to either a polypeptide of the present invention or a known *M. tuberculosis* antigen, such as the 38 kD antigen described above. For example, administration of DNA encoding a polypeptide of the present invention, either "naked" or in a delivery system as described above, may be followed by administration of an antigen in order to enhance the protective immune effect of the vaccine.

Routes and frequency of administration, as well as dosage, will vary from individual to individual and may parallel those currently being used in immunization using BCG. In general, the pharmaceutical compositions and vaccines may be administered by injection (e.g., intracutaneous, intramuscular, intravenous or subcutaneous), intranasally (e.g., by aspiration) or orally. Between 1 and 3 doses may be administered for a 1-36 week period. Preferably, 3 doses are administered, at

intervals of 3-4 months, and booster vaccinations may be given periodically thereafter. Alternate protocols may be appropriate for individual patients. A suitable dose is an amount of polypeptide or DNA that, when administered as described above, is capable of raising an immune response in an immunized patient sufficient to protect the patient  
5 from *M. tuberculosis* infection for at least 1-2 years. In general, the amount of polypeptide present in a dose (or produced *in situ* by the DNA in a dose) ranges from about 1 pg to about 100 mg per kg of host, typically from about 10 pg to about 1 mg, and preferably from about 100 pg to about 1 µg. Suitable dose sizes will vary with the size of the patient, but will typically range from about 0.1 mL to about 5 mL.

10 While any suitable carrier known to those of ordinary skill in the art may be employed in the pharmaceutical compositions of this invention, the type of carrier will vary depending on the mode of administration. For parenteral administration, such as subcutaneous injection, the carrier preferably comprises water, saline, alcohol, a fat, a wax or a buffer. For oral administration, any of the above carriers or a solid carrier,  
15 such as mannitol, lactose, starch, magnesium stearate, sodium saccharine, talcum, cellulose, glucose, sucrose, and magnesium carbonate, may be employed. Biodegradable microspheres (e.g., polylactic galactide) may also be employed as carriers for the pharmaceutical compositions of this invention. Suitable biodegradable microspheres are disclosed, for example, in U.S. Patent Nos. 4,897,268 and 5,075,109.

20 Any of a variety of adjuvants may be employed in the vaccines of this invention to nonspecifically enhance the immune response. Most adjuvants contain a substance designed to protect the antigen from rapid catabolism, such as aluminum hydroxide or mineral oil, and a nonspecific stimulator of immune responses, such as lipid A, *Bordetella pertussis* or *Mycobacterium tuberculosis*. Suitable adjuvants are  
25 commercially available as, for example, Freund's Incomplete Adjuvant and Freund's Complete Adjuvant (Difco Laboratories) and Merck Adjuvant 65 (Merck and Company, Inc., Rahway, NJ). Other suitable adjuvants include alum, biodegradable microspheres, monophosphoryl lipid A and quil A.

In another aspect, this invention provides methods for using one or more  
30 of the polypeptides described above to diagnose tuberculosis using a skin test. As used

herein, a "skin test" is any assay performed directly on a patient in which a delayed-type hypersensitivity (DTH) reaction (such as swelling, reddening or dermatitis) is measured following intradermal injection of one or more polypeptides as described above. Such injection may be achieved using any suitable device sufficient to contact the 5 polypeptide or polypeptides with dermal cells of the patient, such as a tuberculin syringe or 1 mL syringe. Preferably, the reaction is measured at least 48 hours after injection, more preferably 48-72 hours.

The DTH reaction is a cell-mediated immune response, which is greater in patients that have been exposed previously to the test antigen (*i.e.*, the immunogenic 10 portion of the polypeptide employed, or a variant thereof). The response may be measured visually, using a ruler. In general, a response that is greater than about 0.5 cm in diameter, preferably greater than about 1.0 cm in diameter, is a positive response, indicative of tuberculosis infection, which may or may not be manifested as an active disease.

15 The polypeptides of this invention are preferably formulated, for use in a skin test, as pharmaceutical compositions containing a polypeptide and a physiologically acceptable carrier, as described above. Such compositions typically contain one or more of the above polypeptides in an amount ranging from about 1  $\mu$ g to about 100  $\mu$ g, preferably from about 10  $\mu$ g to about 50  $\mu$ g in a volume of 0.1 mL. 20 Preferably, the carrier employed in such pharmaceutical compositions is a saline solution with appropriate preservatives, such as phenol and/or Tween 80<sup>TM</sup>.

In a preferred embodiment, a polypeptide employed in a skin test is of sufficient size such that it remains at the site of injection for the duration of the reaction period. In general, a polypeptide that is at least 9 amino acids in length is sufficient. 25 The polypeptide is also preferably broken down by macrophages within hours of injection to allow presentation to T-cells. Such polypeptides may contain repeats of one or more of the above sequences and/or other immunogenic or nonimmunogenic sequences.

The following Examples are offered by way of illustration and not by way of limitation.

### EXAMPLES

5

#### EXAMPLE 1

##### PURIFICATION AND CHARACTERIZATION OF POLYPEPTIDES FROM *M. TUBERCULOSIS* CULTURE FILTRATE

10 This example illustrates the preparation of *M. tuberculosis* soluble polypeptides from culture filtrate. Unless otherwise noted, all percentages in the following example are weight per volume.

15 *M. tuberculosis* (either H37Ra, ATCC No. 25177, or H37Rv, ATCC No. 25618) was cultured in sterile GAS media at 37°C for fourteen days. The media was then vacuum filtered (leaving the bulk of the cells) through a 0.45 µ filter into a sterile 2.5 L bottle. The media was next filtered through a 0.2 µ filter into a sterile 4 L bottle and NaN<sub>3</sub> was added to the culture filtrate to a concentration of 0.04%. The bottles were then placed in a 4°C cold room.

20 The culture filtrate was concentrated by placing the filtrate in a 12 L reservoir that had been autoclaved and feeding the filtrate into a 400 ml Amicon stir cell which had been rinsed with ethanol and contained a 10,000 kDa MWCO membrane. The pressure was maintained at 60 psi using nitrogen gas. This procedure reduced the 12 L volume to approximately 50 ml.

25 The culture filtrate was dialyzed into 0.1% ammonium bicarbonate using a 8,000 kDa MWCO cellulose ester membrane, with two changes of ammonium bicarbonate solution. Protein concentration was then determined by a commercially available BCA assay (Pierce, Rockford, IL).

30 The dialyzed culture filtrate was then lyophilized, and the polypeptides resuspended in distilled water. The polypeptides were dialyzed against 0.01 mM 1,3 bis[tris(hydroxymethyl)-methylamino]propane, pH 7.5 (Bis-Tris propane buffer), the

initial conditions for anion exchange chromatography. Fractionation was performed using gel profusion chromatography on a POROS 146 II Q/M anion exchange column 4.6 mm x 100 mm (Perseptive BioSystems, Framingham, MA) equilibrated in 0.01 mM Bis-Tris propane buffer pH 7.5. Polypeptides were eluted with a linear 0-0.5 M NaCl 5 gradient in the above buffer system. The column eluent was monitored at a wavelength of 220 nm.

The pools of polypeptides eluting from the ion exchange column were dialyzed against distilled water and lyophilized. The resulting material was dissolved in 0.1% trifluoroacetic acid (TFA) pH 1.9 in water, and the polypeptides were purified on 10 a Delta-Pak C18 column (Waters, Milford, MA) 300 Angstrom pore size, 5 micron particle size (3.9 x 150 mm). The polypeptides were eluted from the column with a linear gradient from 0-60% dilution buffer (0.1% TFA in acetonitrile). The flow rate was 0.75 ml/minute and the HPLC eluent was monitored at 214 nm. Fractions containing the eluted polypeptides were collected to maximize the purity of the 15 individual samples. Approximately 200 purified polypeptides were obtained.

The purified polypeptides were then screened for the ability to induce T-cell proliferation in PBMC preparations. The PBMCs from donors known to be PPD skin test positive and whose T-cells were shown to proliferate in response to PPD and crude soluble proteins from MTB were cultured in medium comprising RPMI 1640 20 supplemented with 10% pooled human serum and 50 µg/ml gentamicin. Purified polypeptides were added in duplicate at concentrations of 0.5 to 10 µg/mL. After six days of culture in 96-well round-bottom plates in a volume of 200 µl, 50 µl of medium was removed from each well for determination of IFN-γ levels, as described below. The plates were then pulsed with 1 µCi/well of tritiated thymidine for a further 18 25 hours, harvested and tritium uptake determined using a gas scintillation counter. Fractions that resulted in proliferation in both replicates three fold greater than the proliferation observed in cells cultured in medium alone were considered positive.

IFN-γ was measured using an enzyme-linked immunosorbent assay (ELISA). ELISA plates were coated with a mouse monoclonal antibody directed to 30 human IFN-γ (PharMingen, San Diego, CA) in PBS for four hours at room temperature.

- Wells were then blocked with PBS containing 5% (W/V) non-fat dried milk for 1 hour at room temperature. The plates were then washed six times in PBS/0.2% TWEEN-20 and samples diluted 1:2 in culture medium in the ELISA plates were incubated overnight at room temperature. The plates were again washed and a polyclonal rabbit
- 5 anti-human IFN- $\gamma$  serum diluted 1:3000 in PBS/10% normal goat serum was added to each well. The plates were then incubated for two hours at room temperature, washed and horseradish peroxidase-coupled anti-rabbit IgG (Sigma Chemical So., St. Louis, MO) was added at a 1:2000 dilution in PBS/5% non-fat dried milk. After a further two hour incubation at room temperature, the plates were washed and TMB substrate added.
- 10 The reaction was stopped after 20 min with 1 N sulfuric acid. Optical density was determined at 450 nm using 570 nm as a reference wavelength. Fractions that resulted in both replicates giving an OD two fold greater than the mean OD from cells cultured in medium alone, plus 3 standard deviations, were considered positive.

For sequencing, the polypeptides were individually dried onto

15 Biobrene™ (Perkin Elmer/Applied BioSystems Division, Foster City, CA) treated glass fiber filters. The filters with polypeptide were loaded onto a Perkin Elmer/Applied BioSystems Division Procise 492 protein sequencer. The polypeptides were sequenced from the amino terminal and using traditional Edman chemistry. The amino acid sequence was determined for each polypeptide by comparing the retention time of the

20 PTH amino acid derivative to the appropriate PTH derivative standards.

Using the procedure described above, antigens having the following N-terminal sequences were isolated:

- (a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Thr-Xaa-Asn-Tyr-Gly-  
Gln-Val-Val-Ala-Ala-Leu; (SEQ ID No. 54)
- 25 (b) Ala-Val-Glu-Ser-Gly-Met-Leu-Ala-Leu-Gly-Thr-Pro-Ala-Pro-  
Ser; (SEQ ID No. 55)
- (c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-  
Ala-Lys-Glu-Gly-Arg; (SEQ ID No. 56)
- (d) Tyr-Tyr-Trp-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Trp-Gly-  
30 Pro; (SEQ ID No. 57)

- (e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val;  
(SEQ ID No. 58)
- (f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro; (SEQ ID  
No. 59)
- 5 (g) Asp-Pro-Glu-Pro-Ala-Pro-Pro-Val-Pro-Thr-Ala-Ala-Ala-  
Pro-Pro-Ala; (SEQ ID No. 60) and
- (h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-  
Gly; (SEQ ID No. 61)

wherein Xaa may be any amino acid.

10 An additional antigen was isolated employing a microbore HPLC purification step in addition to the procedure described above. Specifically, 20 µl of a fraction comprising a mixture of antigens from the chromatographic purification step previously described, was purified on an Aquapore C18 column (Perkin Elmer/Applied Biosystems Division, Foster City, CA) with a 7 micron pore size, column size 1 mm x  
15 100 mm, in a Perkin Elmer/Applied Biosystems Division Model 172 HPLC. Fractions were eluted from the column with a linear gradient of 1%/minute of acetonitrile (containing 0.05% TFA) in water (0.05% TFA) at a flow rate of 80 µl/minute. The eluent was monitored at 250 nm. The original fraction was separated into 4 major peaks plus other smaller components and a polypeptide was obtained which was shown to  
20 have a molecular weight of 12.054 Kd (by mass spectrometry) and the following N-terminal sequence:

25 (i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Gln-  
Thr-Ser-Leu-Leu-Asn-Asn-Leu-Ala-Asp-Pro-Asp-Val-Ser-Phe-  
Ala-Asp (SEQ ID No. 62).

30 This polypeptide was shown to induce proliferation and IFN-γ production in PBMC preparations using the assays described above.

Additional soluble antigens were isolated from *M. tuberculosis* culture filtrate as follows. *M. tuberculosis* culture filtrate was prepared as described above. Following dialysis against Bis-Tris propane buffer, at pH 5.5, fractionation was performed using anion exchange chromatography on a Poros QE column 4.6 x 100 mm

(Perseptive Biosystems) equilibrated in Bis-Tris propane buffer pH 5.5. Polypeptides were eluted with a linear 0-1.5 M NaCl gradient in the above buffer system at a flow rate of 10 ml/min. The column eluent was monitored at a wavelength of 214 nm.

The fractions eluting from the ion exchange column were pooled and 5 subjected to reverse phase chromatography using a Poros R2 column 4.6 x 100 mm (Perseptive Biosystems). Polypeptides were eluted from the column with a linear gradient from 0-100% acetonitrile (0.1% TFA) at a flow rate of 5 ml/min. The eluent was monitored at 214 nm.

Fractions containing the eluted polypeptides were lyophilized and 10 resuspended in 80 µl of aqueous 0.1% TFA and further subjected to reverse phase chromatography on a Vydac C4 column 4.6 x 150 mm (Western Analytical, Temecula, CA) with a linear gradient of 0-100% acetonitrile (0.1% TFA) at a flow rate of 2 ml/min. Eluent was monitored at 214 nm.

The fraction with biological activity was separated into one major peak 15 plus other smaller components. Western blot of this peak onto PVDF membrane revealed three major bands of molecular weights 14 Kd, 20 Kd and 26 Kd. These polypeptides were determined to have the following N-terminal sequences, respectively:

- (j) Xaa-Asp-Ser-Glu-Lys-Ser-Ala-Thr-Ile-Lys-Val-Thr-Asp-Ala-Ser; (SEQ ID No. 134)
- 20 (k) Ala-Gly-Asp-Thr-Xaa-Ile-Tyr-Ile-Val-Gly-Asn-Leu-Thr-Ala-Asp; (SEQ ID No. 135) and
- (l) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Leu-Gly-Gly-Thr-Val-Gln-Ala-Gly; (SEQ ID No. 136), wherein Xaa may be any amino acid.

Using the assays described above, these polypeptides were shown to induce 25 proliferation and IFN- $\gamma$  production in PBMC preparations. Figs. 1A and B show the results of such assays using PBMC preparations from a first and a second donor, respectively.

DNA sequences that encode the antigens designated as (a), (c), (d) and 30 (g) above were obtained by screening a genomic *M. tuberculosis* library using  $^{32}$ P end labeled degenerate oligonucleotides corresponding to the N-terminal sequence and

containing *M. tuberculosis* codon bias. The screen performed using a probe corresponding to antigen (a) above identified a clone having the sequence provided in SEQ ID No. 101. The polypeptide encoded by SEQ ID No. 101 is provided in SEQ ID No. 102. The screen performed using a probe corresponding to antigen (g) above 5 identified a clone having the sequence provided in SEQ ID No. 52. The polypeptide encoded by SEQ ID No. 52 is provided in SEQ ID No. 53. The screen performed using a probe corresponding to antigen (d) above identified a clone having the sequence provided in SEQ ID No. 24, and the screen performed with a probe corresponding to antigen (c) identified a clone having the sequence provided in SEQ ID No. 25.

10 The above amino acid sequences were compared to known amino acid sequences in the gene bank using the DNA STAR system. The database searched contains some 173,000 proteins and is a combination of the Swiss, PIR databases along with translated protein sequences (Version 87). No significant homologies to the amino acid sequences for antigens (a)-(h) and (l) were detected.

15 The amino acid sequence for antigen (i) was found to be homologous to a sequence from *M. leprae*. The full length *M. leprae* sequence was amplified from genomic DNA using the sequence obtained from GENBANK. This sequence was then used to screen the *M. tuberculosis* library described below in Example 2 and a full length copy of the *M. tuberculosis* homologue was obtained (SEQ ID No. 99).

20 The amino acid sequence for antigen (j) was found to be homologous to a known *M. tuberculosis* protein translated from a DNA sequence. To the best of the inventors' knowledge, this protein has not been previously shown to possess T-cell stimulatory activity. The amino acid sequence for antigen (k) was found to be related to a sequence from *M. leprae*.

25 In the proliferation and IFN- $\gamma$  assays described above, using three PPD positive donors, the results for representative antigens provided above are presented in Table 1:

TABLE 1  
RESULTS OF PBMC PROLIFERATION AND IFN- $\gamma$  ASSAYS

Sequence	Proliferation	IFN- $\gamma$
(a)	+	-
(c)	+++	+++
(d)	++	++
(g)	+++	+++
(h)	+++	+++

- 5        In Table 1, responses that gave a stimulation index (SI) of between 2 and 4 (compared to cells cultured in medium alone) were scored as +, an SI of 4-8 or 2-4 at a concentration of 1  $\mu$ g or less was scored as ++ and an SI of greater than 8 was scored as +++. The antigen of sequence (i) was found to have a high SI (++) for one donor and lower SI (++ and +) for the two other donors in both proliferation and IFN- $\gamma$  assays.
- 10      These results indicate that these antigens are capable of inducing proliferation and/or interferon- $\gamma$  production.

EXAMPLE 2

USE OF PATIENT SERA TO ISOLATE *M. TUBERCULOSIS* ANTIGENS

- 15      This example illustrates the isolation of antigens from *M. tuberculosis* lysate by screening with serum from *M. tuberculosis*-infected individuals.
- 16      Dessicated *M. tuberculosis* H37Ra (Difco Laboratories) was added to a 2% NP40 solution, and alternately homogenized and sonicated three times. The resulting suspension was centrifuged at 13,000 rpm in microfuge tubes and the supernatant put through a 0.2 micron syringe filter. The filtrate was bound to Macro Prep DEAE beads (BioRad, Hercules, CA). The beads were extensively washed with 20 mM Tris pH 7.5 and bound proteins eluted with 1M NaCl. The 1M NaCl elute was dialyzed overnight against 10 mM Tris, pH 7.5. Dialyzed solution was treated with

- DNase and RNase at 0.05 mg/ml for 30 min. at room temperature and then with  $\alpha$ -D-mannosidase, 0.5 U/mg at pH 4.5 for 3-4 hours at room temperature. After returning to pH 7.5, the material was fractionated via FPLC over a Bio Scale-Q-20 column (BioRad). Fractions were combined into nine pools, concentrated in a Centriprep 10  
5 (Amicon, Beverley, MA) and then screened by Western blot for serological activity using a serum pool from *M. tuberculosis*-infected patients which was not immunoreactive with other antigens of the present invention.

The most reactive fraction was run in SDS-PAGE and transferred to PVDF. A band at approximately 85 Kd was cut out yielding the sequence:

- 10 (m) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-  
Ile-Asn-Val-His-Leu-Val: (SEQ ID No. 137), wherein Xaa may  
be any amino acid.

Comparison of this sequence with those in the gene bank as described above, revealed no significant homologies to known sequences.

15

#### EXAMPLE 3

##### PREPARATION OF DNA SEQUENCES ENCODING *M. TUBERCULOSIS* ANTIGENS

- This example illustrates the preparation of DNA sequences encoding  
20 *M. tuberculosis* antigens by screening a *M. tuberculosis* expression library with sera obtained from patients infected with *M. tuberculosis*, or with anti-sera raised against soluble *M. tuberculosis* antigens.

- 25 A. PREPARATION OF *M. TUBERCULOSIS* SOLUBLE ANTIGENS USING RABBIT ANTI-SERA

- Genomic DNA was isolated from the *M. tuberculosis* strain H37Ra. The DNA was randomly sheared and used to construct an expression library using the Lambda ZAP expression system (Stratagene, La Jolla, CA). Rabbit anti-sera was generated against secretory proteins of the *M. tuberculosis* strains H37Ra, H37Rv and  
30 Erdman by immunizing a rabbit with concentrated supernatant of the *M. tuberculosis* cultures. Specifically, the rabbit was first immunized subcutaneously with 200  $\mu$ g of

protein antigen in a total volume of 2 ml containing 10 µg muramyl dipeptide (Calbiochem, La Jolla, CA) and 1 ml of incomplete Freund's adjuvant. Four weeks later the rabbit was boosted subcutaneously with 100 µg antigen in incomplete Freund's adjuvant. Finally, the rabbit was immunized intravenously four weeks later with 50 µg 5 protein antigen. The anti-sera were used to screen the expression library as described in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989. Bacteriophage plaques expressing immunoreactive antigens were purified. Phagemid from the plaques was rescued and the nucleotide sequences of the *M. tuberculosis* clones deduced.

10       Thirty two clones were purified. Of these, 25 represent sequences that have not been previously identified in human *M. tuberculosis*. Recombinant antigens were expressed and purified antigens used in the immunological analysis described in Example 1. Proteins were induced by IPTG and purified by gel elution, as described in Skeiky et al., *J. Exp. Med.* 181:1527-1537, 1995. Representative sequences of DNA 15 molecules identified in this screen are provided in SEQ ID Nos.: 1-25. The corresponding predicted amino acid sequences are shown in SEQ ID Nos. 63-87.

On comparison of these sequences with known sequences in the gene bank using the databases described above, it was found that the clones referred to hereinafter as TbRA2A, TbRA16, TbRA18, and TbRA29 (SEQ ID Nos. 76, 68, 70, 75) 20 show some homology to sequences previously identified in *Mycobacterium leprae* but not in *M. tuberculosis*. TbRA11, TbRA26, TbRA28 and TbDPEP (SEQ ID Nos.: 65, 73, 74, 53) have been previously identified in *M. tuberculosis*. No significant homologies were found to TbRA1, TbRA3, TbRA4, TbRA9, TbRA10, TbRA13, TbRA17, TbRa19, TbRA29, TbRA32, TbRA36 and the overlapping clones TbRA35 25 and TbRA12 (SEQ ID Nos. 63, 77, 81, 82, 64, 67, 69, 71, 75, 78, 80, 79, 66). The clone TbRa24 is overlapping with clone TbRa29.

The results of PBMC proliferation and interferon- $\gamma$  assays performed on representative recombinant antigens, and using T-cell preparations from several different *M. tuberculosis*-immune patients, are presented in Tables 2 and 3, 30 respectively.

**TABLE 2**  
**RESULTS OF PBMC PROLIFERATION TO REPRESENTATIVE SOLUBLE ANTIGENS**

Antigen	Patient												
	1	2	3	4	5	6	7	8	9	10	11	12	
TbRa1	-	-	±	++	-	-	±	±	-	-	+	±	-
TbRa3	-	±	++	-	±	-	-	++	±	-	-	-	-
TbRa9	-	-	nt	nt	++	nt							
TbRa10	-	-	±	±	+	nt	±	-	+	±	±	-	-
TbRa11	±	±	+	++	+	+	nt	-	++	++	++	±	nt
TbRa12	-	-	+	+	±	++	+	±	±	-	+	-	-
TbRa16	nt	nt	nt	-	+	nt							
TbRa24	nt	nt	nt	-	-	nt							
TbRa26	-	+	nt	nt	-	-	nt						
TbRa29	nt	nt	nt	nt	-	-	nt						
TbRa35	++	nt	++	++	++	nt	++	++	++	++	++	++	nt
TbRaB	nt	nt	nt	nt	-	-	nt						
TbRaC	nt	nt	nt	nt	-	-	nt						
TbRaD	nt	nt	nt	nt	-	-	nt						
AAMK	-	-	±	-	-	nt	-	-	-	nt	±	nt	nt
YY	-	-	-	-	-	-	nt	-	-	nt	+	nt	nt
DPEP	-	+	-	++	-	-	nt	++	±	+	±	±	nt
Control	-	-	-	-	-	-	-	-	-	-	-	-	-

nt = not tested

**TABLE 3**  
**RESULTS OF PBMC INTERFERON-γ PRODUCTION TO REPRESENTATIVE SOLUBLE ANTIGENS**

In Tables 2 and 3, responses that gave a stimulation index (SI) of between 1.2 and 2 (compared to cells cultured in medium alone) were scored as  $\pm$ , a SI of 2-4 was scored as +, as SI of 4-8 or 2-4 at a concentration of 1  $\mu$ g or less was scored as ++ and an SI of greater than 8 was scored as +++. In addition, the effect of 5 concentration on proliferation and interferon- $\gamma$  production is shown for two of the above antigens in the attached Figure. For both proliferation and interferon- $\gamma$  production, TbRa3 was scored as ++ and TbRa9 as +.

These results indicate that these soluble antigens can induce proliferation and/or interferon- $\gamma$  production in T-cells derived from an *M. tuberculosis*-immune 10 individual.

B. USE OF PATIENT SERA TO IDENTIFY DNA SEQUENCES ENCODING M. TUBERCULOSIS ANTIGENS

The genomic DNA library described above, and an additional H37Rv 15 library, were screened using pools of sera obtained from patients with active tuberculosis. To prepare the H37Rv library, *M. tuberculosis* strain H37Rv genomic DNA was isolated, subjected to partial *Sau3A* digestion and used to construct an expression library using the Lambda Zap expression system (Stratagene, La Jolla, Ca). Three different pools of sera, each containing sera obtained from three individuals with 20 active pulmonary or pleural disease, were used in the expression screening. The pools were designated TbL, TbM and TbH, referring to relative reactivity with H37Ra lysate (*i.e.*, TbL = low reactivity, TbM = medium reactivity and TbH = high reactivity) in both ELISA and immunoblot format. A fourth pool of sera from seven patients with active pulmonary tuberculosis was also employed. All of the sera lacked increased reactivity 25 with the recombinant 38 kD *M. tuberculosis* H37Ra phosphate-binding protein.

All pools were pre-adsorbed with *E. coli* lysate and used to screen the H37Ra and H37Rv expression libraries, as described in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989. Bacteriophage plaques expressing immunoreactive antigens were purified. 30 Phagemid from the plaques was rescued and the nucleotide sequences of the *M. tuberculosis* clones deduced.

Thirty two clones were purified. Of these, 31 represented sequences that had not been previously identified in human *M. tuberculosis*. Representative sequences of the DNA molecules identified are provided in SEQ ID Nos.: 26-51 and 105. Of these, TbH-8 and TbH-8-2 (SEQ. ID NO. 105) are non-contiguous DNA sequences 5 from the same clone, and TbH-4 (SEQ. ID NO. 43) and TbH-4-FWD (SEQ. ID NO. 44) are non-contiguous sequences from the same clone. Amino acid sequences for the antigens hereinafter identified as Tb38-1, TbH-4, TbH-8, TbH-9, and TbH-12 are shown in SEQ ID Nos.: 88-92. Comparison of these sequences with known sequences 10 in the gene bank using the databases identified above revealed no significant homologies to TbH-4, TbH-8, TbH-9 and TbM-3, although weak homologies were found to TbH-9. TbH-12 was found to be homologous to a 34 kD antigenic protein 15 previously identified in *M. paratuberculosis* (Acc. No. S28515). Tb38-1 was found to be located 34 base pairs upstream of the open reading frame for the antigen ESAT-6 previously identified in *M. bovis* (Acc. No. U34848) and in *M. tuberculosis* (Sorensen et al., *Infec. Immun.* 63:1710-1717, 1995).

Probes derived from Tb38-1 and TbH-9, both isolated from an H37Ra library, were used to identify clones in an H37Rv library. Tb38-1 hybridized to Tb38-1F2, Tb38-1F3, Tb38-1F5 and Tb38-1F6 (SEQ. ID NOS. 112, 113, 116, 118, and 119). (SEQ ID NOS. 112 and 113 are non-contiguous sequences from clone Tb38-20 1F2.) Two open reading frames were deduced in Tb38-IF2; one corresponds to Tb37FL (SEQ. ID. NO. 114), the second, a partial sequence, may be the homologue of Tb38-1 and is called Tb38-IN (SEQ. ID NO. 115). The deduced amino acid sequence of Tb38-1F3 is presented in SEQ. ID. NO. 117. A TbH-9 probe identified three clones in the H37Rv library: TbH-9-FL (SEQ. ID NO. 106), which may be the homologue of TbH-9 25 (R37Ra), TbH-9-1 (SEQ. ID NO. 108), and TbH-9-4 (SEQ. ID NO. 110), all of which are highly related sequences to TbH-9. The deduced amino acid sequences for these three clones are presented in SEQ ID NOS. 107, 109 and 111.

The results of T-cell assays performed on Tb38-1, ESAT-6 and other representative recombinant antigens are presented in Tables 4A, B and 5, respectively, 30 below:

TABLE 4A  
RESULTS OF PBMC PROLIFERATION TO REPRESENTATIVE ANTIGENS

Antigen	Donor										
	1	2	3	4	5	6	7	8	9	10	11
Tb38.1	+++	+	-	-	-	++	-	+	-	++	+++
ESAT-6	+++	+	+	+	-	+	-	+	+	++	+++
TbH-9	++	++	-	++	±	±	++	++	++	++	++

5

TABLE 4B  
RESULTS OF PBMC INTERFERON-γ PRODUCTION TO REPRESENTATIVE ANTIGENS

Antigen	Donor										
	1	2	3	4	5	6	7	8	9	10	11
Tb38.1	+++	+	-	+	+	+++	-	++	-	+++	+++
ESAT-6	+++	+	+	+	+-	+	-	+	+	+++	+++
TbH-9	++	++	-	+++	±	±	+++	+++	++	+++	++

10

TABLE 5  
SUMMARY OF T-CELL RESPONSES TO REPRESENTATIVE ANTIGENS

Antigen	Proliferation			Interferon- $\gamma$			total
	patient 4	patient 5	patient 6	patient 4	patient 5	patient 6	
TbH9	++	++	++	+++	++	++	13
TbM7	-	+	-	++	+	-	4
TbH5	-	+	+	++	++	++	8
TbL23	-	+	±	++	++	+	7.5
TbH4	-	++	±	++	++	±	7
- control	-	-	-	-	-	-	0

5

These results indicate that both the inventive *M. tuberculosis* antigens and ESAT-6 can induce proliferation and/or interferon- $\gamma$  production in T-cells derived from an *M. tuberculosis*-immune individual. To the best of the inventors' knowledge, ESAT-6 has not been previously shown to stimulate human immune responses

- 10 A set of six overlapping peptides covering the amino acid sequence of the antigen Tb38-1 was constructed using the method described in Example 4. The sequences of these peptides, hereinafter referred to as pep1-6, are provided in SEQ ID Nos. 93-98, respectively. The results of T-cell assays using these peptides are shown in Tables 6 and 7. These results confirm the existence, and help to localize T-cell epitopes  
15 within Tb38-1 capable of inducing proliferation and interferon- $\gamma$  production in T-cells derived from an *M. tuberculosis* immune individual.

TABLE 6  
RESULTS OF PBMC PROLIFERATION TO Tb38-1 PEPTIDES

**TABLE 7**  
**RESULTS OF PBMC INTERFERON- $\gamma$  PRODUCTION TO TB38-1 PEPTIDES**

EXAMPLE 4PURIFICATION AND CHARACTERIZATION OF A POLYPEPTIDE FROM TUBERCULIN PURIFIED PROTEIN DERIVATIVE

5

An *M. tuberculosis* polypeptide was isolated from tuberculin purified protein derivative (PPD) as follows.

PPD was prepared as published with some modification (Seibert, F. et al., Tuberculin purified protein derivative. Preparation and analyses of a large quantity 10 for standard. The American Review of Tuberculosis 44:9-25, 1941).

*M. tuberculosis* Rv strain was grown for 6 weeks in synthetic medium in roller bottles at 37°C. Bottles containing the bacterial growth were then heated to 100° C in water vapor for 3 hours. Cultures were sterile filtered using a 0.22 µ filter and the liquid phase was concentrated 20 times using a 3 kD cut-off membrane. Proteins were 15 precipitated once with 50% ammonium sulfate solution and eight times with 25% ammonium sulfate solution. The resulting proteins (PPD) were fractionated by reverse phase liquid chromatography (RP-HPLC) using a C18 column (7.8 x 300 mM; Waters, Milford, MA) in a Biocad HPLC system (Perseptive Biosystems, Framingham, MA). Fractions were eluted from the column with a linear gradient from 0-100% buffer (0.1% 20 TFA in acetonitrile). The flow rate was 10 ml/minute and eluent was monitored at 214 nm and 280 nm.

Six fractions were collected, dried, suspended in PBS and tested individually in *M. tuberculosis*-infected guinea pigs for induction of delayed type hypersensitivity (DTH) reaction. One fraction was found to induce a strong DTH 25 reaction and was subsequently fractionated further by RP-HPLC on a microbore Vydac C18 column (Cat. No. 218TP5115) in a Perkin Elmer/Applied Biosystems Division Model 172 HPLC. Fractions were eluted with a linear gradient from 5-100% buffer (0.05% TFA in acetonitrile) with a flow rate of 80 µl/minute. Eluent was monitored at 215 nm. Eight fractions were collected and tested for induction of DTH in *M. 30 tuberculosis*-infected guinea pigs. One fraction was found to induce strong DTH of

about 16 mm induration. The other fractions did not induce detectable DTH. The positive fraction was submitted to SDS-PAGE gel electrophoresis and found to contain a single protein band of approximately 12 kD molecular weight.

This polypeptide, herein after referred to as DPPD, was sequenced from 5 the amino terminal using a Perkin Elmer/Applied Biosystems Division Procise 492 protein sequencer as described above and found to have the N-terminal sequence shown in SEQ ID No.: 129. Comparison of this sequence with known sequences in the gene bank as described above revealed no known homologies. Four cyanogen bromide fragments of DPPD were isolated and found to have the sequences shown in SEQ ID 10 Nos.: 130-133.

The ability of the antigen DPPD to stimulate human PBMC to proliferate and to produce IFN- $\gamma$  was assayed as described in Example 1. As shown in Table 8, DPPD was found to stimulate proliferation and elicit production of large quantities of IFN- $\gamma$ ; more than that elicited by commercial PPD.

15

TABLE 8  
RESULTS OF PROLIFERATION AND INTERFERON- $\gamma$  ASSAYS TO DPPD

PBMC Donor	Stimulator	Proliferation (CPM)	IFN- $\gamma$ (OD <sub>450</sub> )
A	Medium	1,089	0.17
	PPD (commercial)	8,394	1.29
	DPPD	13,451	2.21
B	Medium	450	0.09
	PPD (commercial)	3,929	1.26
	DPPD	6,184	1.49
C	Medium	541	0.11
	PPD (commercial)	8,907	0.76
	DPPD	23,024	>2.70

EXAMPLE 5SYNTHESIS OF SYNTHETIC POLYPEPTIDES

5

- Polypeptides may be synthesized on a Millipore 9050 peptide synthesizer using FMOC chemistry with HPTU (O-Benzotriazole-N,N,N',N'-tetramethyluronium hexafluorophosphate) activation. A Gly-Cys-Gly sequence may be attached to the amino terminus of the peptide to provide a method of conjugation or  
10 labeling of the peptide. Cleavage of the peptides from the solid support may be carried out using the following cleavage mixture: trifluoroacetic acid:ethanedithiol:thioanisole:water:phenol (40:1:2:2:3). After cleaving for 2 hours, the peptides may be precipitated in cold methyl-t-butyl-ether. The peptide pellets may then be dissolved in water containing 0.1% trifluoroacetic acid (TFA) and lyophilized prior  
15 to purification by C18 reverse phase HPLC. A gradient of 0%-60% acetonitrile (containing 0.1% TFA) in water (containing 0.1% TFA) may be used to elute the peptides. Following lyophilization of the pure fractions, the peptides may be characterized using electrospray mass spectrometry and by amino acid analysis.
- 20 From the foregoing, it will be appreciated that, although specific embodiments of the invention have been described herein for the purpose of illustration, various modifications may be made without deviating from the spirit and scope of the invention.

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANTS: Corixa Corporation
- (ii) TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY AND DIAGNOSIS OF TUBERCULOSIS
- (iii) NUMBER OF SEQUENCES: 137
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: SEED and BERRY LLP
  - (B) STREET: 6300 Columbia Center, 701 Fifth Avenue
  - (C) CITY: Seattle
  - (D) STATE: Washington
  - (E) COUNTRY: USA
  - (F) ZIP: 98104-7092
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0. Version #1.30
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER:
  - (B) FILING DATE: 27-AUG-1996
  - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Maki, David J.
  - (B) REGISTRATION NUMBER: 31.392
  - (C) REFERENCE/DOCKET NUMBER: 210121.411PC
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: (206) 622-4900
  - (B) TELEFAX: (206) 682-6031

## (2) INFORMATION FOR SEQ ID NO:1:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 766 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CGAGGCACCG GTAGTTGAA CCAAACGCAC AATCGACGGG CAAACGAACG GAAGAACACA	60
ACCATGAAAGA TGGTGAAATC GATGCCGCA GGTCTGACCG CCGCGGCTGC AATCGGCAC	120
GCTGCGGCCG GTGTGACTTC GATCATGGCT GGCGGCCCGG TCGTATAACCA GATGCAGCCG	180
GTCGTCTTCG GCGGCCACT GCCGTTGGAC CCGGCATCCG CCCCTGACGT CCCGACCGCC	240
GCCCAGTTGA CCAGCCTGCT CAACAGCCTC GCCGATCCA ACGTGTGTT TGCGAACAAAG	300
GGCAGTCTGG TCGAGGGCGG CATCGGGGC ACCGAGGCAC GCATGCCGA CCACAAGCTG	360
AAGAAGGCCG CCGAGCACGG GGATCTGCCG CTGTCGTTCA GCGTGACGAA CATCCAGCCG	420
GCGGCCGCG GTTCGGCCAC CGCCGACGTT TCCGTCTCGG GTCCGAAGCT CTCGTCGCCG	480
GTCACGCAGA ACGTCACGTT CGTGAATCAA GGCGGCTGGA TGCTGTCACG CGCATCGGCCG	540
ATGGAGTTGC TGCAGGCCGC AGGGNAACTG ATTGGCGGGC CGGNTTCAGC CCGCTGTTCA	600
GCTACGCCGC CCGCCTGGTG ACGCGTCCAT GTCGAACACT CGCGCGTGTGTA GCACGGTGCG	660
GTNTGCGCAG GGNCGCACGC ACCGCCCCGT GCAAGCCGTC CTCGAGATAG GTGGTGNCTC	720
GNCACCAGNG ANCACCCCCN NNTCGNCNN TCTCGNTGNT GNATGA	766

## (2) INFORMATION FOR SEQ ID NO:2:

- (A) LENGTH: 752 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

ATGCATCACCA ATCACCATCA CGATGAAGTC ACGGTAGAGA CGACCTCCGT CTTCCGCGCA	60
GACTTCCTCA GCGAGCTGGA CGCTCCTGCG CAAGCAGGGTA CGGAGAGCGC GGTCTCCGGG	120
GTGGAAGGGC TCCCGCCGGG CTCGGCGTTG CTGGTAGTCA AACGAGGCCA CAACGCCGGG	180
TCCCAGGTTCC TACTCGACCA AGCCATCACG TCGGCTGGTC GGCACTCCGA CAGCGACATA	240
TTTCTCGACG ACGTGACCGT GAGCCGTCGC CATGCTGAAT TCCGGTTGGA AAACAACGAA	300
TTCAATGTCG TCGATGTCGG GAGTCTAAC GGCACCTACG TCAACCGCGA GCCCCGTGGAT	360
TCGGCGGTGC TGGCGAACGG CGACGAGGTC CAGATCGGCA AGCTCCGGTT GGTGTTCTTG	420
ACCGGACCCA AGCAAGGCGA GGATGACGGG AGTACCGGGG GCCCGTGAGC GCACCCGATA	480
GCCCCCGCGCT GGCCGGGATG TCGATCGGGG CGGTCTCCG ACCTGCTACG ACCGGATTIT	540
CCCTGATGTC CACCATCTCC AAGATTGAT TCTTGGGAGG CTTGAGGGTC NGGGTGACCC	600
CCCCGCGGGC CTCATTNGG GGTNTCGGCN GGTTTCACCC CNTACCNACT GCCNCCCGN	660
TTGCNAATTG NTTCTTCNCT GCCCNAAAAG GGACCNNTAN CTTGCCGCTN GAAANGTNA	720
TCCNGGGCCC NTCCTNGAAN CCCCNCCCC CT	752

## (2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 813 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CATATGCATC ACCATCACCA TCACACTTCT AACCGCCCAG CGCGTCGGGG GCGTCGAGCA	60
CCACGCGACA CCGGGCCCCGA TCGATCTGCT AGCTTGAGTC TGGTCAGGCA TCGTCGTAG	120
CAGCGCGATG CCCTATGTTT GTCGTCGACT CAGATATCGC GGCAATCCAA TCTCCCGCCT	180
GC GGCCGGCG GTGCTGCAA CTACTCCGG AGGAATTTCG ACGTGCAT CAAGATCTTC	240
ATGCTGGTCA CGGCTGTCGT TTTGCTCTGT TGTTGGGTG TGGCCACGGC CGCGCCCAAG	300
ACCTACTGCG AGGAGTTGAA AGGCACCGAT ACCGGCCAGG CGTGCAGAT TCAAATGTCC	360
GACCCGGCCT ACAACATCAA CATCAGCCTG CCCAGTTACT ACCCCGACCA GAAGTCGCTG	420
GAAAATTACA TCGCCCAGAC GCGCGACAAG TTCCCTCAGCG CGGCCACATC GTCCACTCCA	480
CGCGAAGCCC CCTACGAATT GAATATCACC TCGGCCACAT ACCAGTCCGC GATAACGCCG	540
CGTGGTACGC AGGCCGTGGT GCTCAMGGTC TACCACAACG CCGGCGGCAC GCACCCAACG	600
ACCACGTACA AGGCCTTCGA TTGGGACCAAG GCCTATCGCA AGCCAATCAC CTATGACACG	660
CTGTGGCAGG CTGACACCGA TCCGCTGCCA GTCGTCTTCC CCATTGTTGC AAGGTGAAC	720
GAGCAACGCA GACCGGGACA ACWGGTATCG ATAGCCGCN AATGCCGGCT TGGAACCCNG	780
TGAAATTATC ACAACTTCGC AGTCACNAAA NAA	813

## (2) INFORMATION FOR SEQ ID NO:4:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 447 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CGGTATGAAC ACGGCCGCGT CCGATAACTT CCAGCTGTCC	CAGGGTGGGC AGGGATT CGC	60
CATTCCGATC GGGCAGGCGA TGGCGATCGC	GGGCCAGATC CGATCGGTG	GGGGGTCA
CCCCGTTCAT ATCGGGCTA CGCCCTTACT CGGCTTGGT	GTTGTCGACA ACAACGGCAA	180
CGGCGCACGA GTGCAACGCG TGGTCGGGAG CGCTCCGGCG	GCAAGTCTCG GCATCTCAC	240
GGCGACGTG ATCACCGCGG TGACCGGCGC TCCGATCAAC	TCGGCCACCG CGATGGCGGA	300
GGCGCTTAAC GGGCATCATC CGGGTGAAGT	CATCTCGGTG AACTGGCAAA CCAAGTCGGG	360
CGGCACGCGT ACAGGGAACG TGACATTGGC CGAGGGACCC	CCGGCCTGAT TTGTCGYGG	420
ATACCACCCG CGGGCCGGCC AATTGGA		447

## (2) INFORMATION FOR SEQ ID NO:5:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 604 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GTCCCAC TGC GGTCGCCGAG TATGTCGCC AGCAAATGTC	TGGCAGCCGC CCAACGGAAT	60
CCGGTGATCC GACGTCGCAG GTTGTGAAAC CCGCCGCCGC	GGAAGTATCG GTCCATGCCT	120
AGCCCCGGCGA CGGGCAGCGC CGGAATGGCG CGAGTGAGGA	GGCAGGGCAAT TTGGCGGGGC	180
CCGGCGACGG NGAGCGCCGG AATGGCGCGA GTGAGGAGGT	GGNCAGTCAT GCCCAGNGTG	240
ATCCAATCAA CCTGNATTG GNCTGNGGGN CCATTGACA ATCGAGGTAG	TGAGCGCAA	300
TGAATGATGG AAAACGGGNG GNACGTCCG NTGTTCTGGT	GGTGNTAGGT GNCTGNCTGG	360

NGTNGNGGNT ATCAGGATGT TCTTCGNCGA AANCTGATGN CGAGGAACAG GGTGTNCCCG	420
NNANNCCNAN GGNGTCCNAN CCCNNNNTCC TCGNCGANAT CANANAGNCG NTTGATGNGA	480
NAAAAGGGTG GANCAGNNNN AANTNGNGGN CCNAANAANC NNNANNGNNG NNAGNTNGNT	540
NNNTNTNNC ANNNNNNNTG NNGNNGNNCN NNNCAANCNN NTNNNNNGNAA NNGGNTTNTT	600
NAAT	604

## (2) INFORMATION FOR SEQ ID NO:6:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 633 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

TTGCANGTCG AACCACCTCA CTAAGGGAA CAAAAGCTNG AGCTCCACCG CGGTGGCGGC	60
CGCTCTAGAA CTAGTGKATM YYYCKGGCTG CAGSAATYCG GYACGAGCAT TAGGACAGTC	120
TAACGGTCCT GTTACGGTGA TCGAATGACC GACGACATCC TGCTGATCGA CACCGACGAA	180
CGGGTGCAGAA CCCTCACCCCT CAACCGGCCG CAGTCCCGYA ACGCGCTCTC GGCGGGCGCTA	240
CGGGATCGGT TTTTCGCGGY GTTGGYCGAC GCCGAGGYCG ACGACGACAT CGACGTCGTC	300
ATCCTCACCG GYGCCGATCC GGTGTTCTGC GCCGGACTGG ACCTCAAGGT AGCTGGCCGG	360
GCAGACCGCG CTGCCGGACA TCTCACCGCG GTGGGCGGCC ATGACCAAGC CGGTGATCGG	420
CGCGATCAAC GGCGCCGCGG TCACCGGCCG GCTCGAACTG GCGCTGTACT GCGACATCCT	480
GATCGCCTCC GAGCACGCCG GCTTCGNCGA CACCCACGCC CGGGTGGGGC TGCTGCCAC	540
CTGGGGACTC AGTGTGTGCT TGCCGAAAA GGTCGGCATC GGNCTGGGCC GGTGGATGAG	600
CCTGACCGGC GACTACCTGT CCGTGACCGA CGC	633

## (2) INFORMATION FOR SEQ ID NO:7:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1362 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CGACGACGAC GGCGCCGGAG AGCGGGCGCG AACGGCGATC GACGCCGGCCC TGGCCAGAGT	60
CGGCACCACC CAGGAGGGAG TCGAATCATG AAATTTGTCA ACCATATTGA GCCCGTCGCG	120
CCCCGCCGAG CGGGCGGCGC GGTCGCCGAG GTCTATGCCG AGGCCCGCCG CGAGTTCGGC	180
CGGCTGCCCG AGCCGCTCGC CATGCTGTCC CCGGACGAGG GACTGCTCAC CGCCGGCTGG	240
GCGACGTTGC GCGAGACACT GCTGGTGGGC CAGGTGCCGC GTGGCCGCAA GGAAGCCGTC	300
GCCGCCGCCG TCGCGGCCAG CCTGCGCTGC CCCTGGTGCG TCGACGCACA CACCACCATG	360
CTGTACGCGG CAGGCCAAAC CGACACCGCC GCGGCGATCT TGGCCGGCAC AGCACCTGCC	420
GCCGGTGACC CGAACGCGCC GTATGTGGCG TGGGCCGAG GAACCGGGAC ACCGGCGGGA	480
CCGCCGGCAC CGTCGGCCC GGATGTCGCC GCCGAATACC TGGGCACCGC GGTGCAATTG	540
CACTTCATCG CACGCCCTGGT CCTGGTGCTG CTGGACGAAA CCTTCCTGCC GGGGGGCCG	600
CGCGCCCCAAC AGCTCATGCG CCGCGCCGGT GGACTGGTGT TCGCCCGCAA GGTGCGCGC	660
GAGCATCGGC CGGGCCGCTC CACCCGCCGG CTCGAGCCGC GAACGCTGCC CGACGATCTG	720
GCATGGGCAA CACCGTCCGA GCCCATAGCA ACCGCGTTCG CCGCGCTCAG CCACCACTG	780
GACACCGCGC CGCACCTGCC GCCACCGACT CGTCAGGTGG TCAGGCGGGT CGTGGGGTCG	840
TGGCACGGCG AGCCAATGCC GATGAGGAGT CGCTGGACGA ACGAGCACAC CGCCGAGCTG	900

CCCGCCGACC TGCACGCGCC CACCCGTCTT GCCCTGCTGA CCGGCCTGGC CCCGCATCAG	960
GTGACCGACG ACGACGTCGC CGCGGCCCGA TCCCTGCTCG ACACCGATGC GGCGCTGGTT	1020
GGCGCCCTGG CCTGGGCCGC CTTCACCGCC GCGCGCGCA TCGGCACCTG GATCGGCGCC	1080
GCCGCCGAGG GCCAGGTGTC GCGGCAAAAC CCGACTGGGT GAGTGTGCGC GCCCTGTCGG	1140
TAGGGTGTCA TCGCTGGCCC GAGGGATCTC GCAGCGCGA ACGGAGGTGG CGACACAGGT	1200
GGAAGCTGCG CCCACTGGCT TGCGCCCCAA CGCCGTGCGT GGCGTTGGT TGGCCGCACT	1260
GGCCGATCAG GTCGGCGCCG GCCCTTGCC GAAGGTCCAG CTAACGTGC CGTCACCGAA	1320
GGACCGGACG GTCACCGGGG GTCACCCCTGC GCGCCCAAGG AA	1362

## (2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1458 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GCGACGACCC CGATATGCCG GGCACCGTAG CGAAAGCCGT CGCCGACGCA CTCGGGGCG	60
GTATCGCTCC CGTTGAGGAC ATTCAAGGACT GCGTGGAGGC CCGGCTGGGG GAAGCCGGTC	120
TGGATGACGT GGCCCGTGT TACATCATCT ACCGGCAGCG GCGCGCCGAG CTGCGGACGG	180
CTAAGGCCTT GCTCGCGTG CGGGACGAGT TAAAGCTGAG CTTGGCGGCC GTGACGGTAC	240
TGCGCGAGCG CTATCTGCTG CACGACGAGC AGGGCCGGCC GGCCGAGTCG ACCGGCGAGC	300
TGATGGACCG ATCGGCGCGC TGTGTCGCGG CGGCCGAGGA CCAGTATGAG CCGGGCTCGT	360
CGAGGGCGGTG GGCCGAGCGG TTG GCCACGC TATTACGCAA CCTGGAATTCTGCCGAATT	420
CGCCCACGTT GATGAACCTCT GGCACCGACC TGGGACTGCT CGCCGGCTGT TTTGTTCTGC	480

CGATTGAGGA TTGCTGCAA TCGATCTTG CGACGCTGGG ACAGGCCGCC GAGCTGCAGC	540
GGGCTGGAGG CGGCACCGGA TATGCGTTCA GCCACCTGCG ACCCGCCGGG GATCGGGTGG	600
CCTCCACGGG CGGCACGGCC AGCGGACCGG TGTCGTTCT ACGGCTGTAT GACAGTGCAG	660
CGGGTGTGGT CTCCATGGGC GGTCGCCGGC GTGGCGCTG TATGGCTGTG CTTGATGTGT	720
CGCACCCGGA TATCTGTGAT TTCGTACCG CCAAGGCCGA ATCCCCCAGC GAGCTCCGC	780
ATTTAACCT ATCGGTTGGT GTGACCGACG CGTTCTGCG GGCGTCGAA CGAACGGCC	840
TACACCGGCT GGTCAATCCG CGAACCGGCA AGATCGTCGC GCGGATGCC GCGGCCGAGC	900
TGTTCGACGC CATCTGAAA GCCGCGCACG CCGGTGGCGA TCCCAGGCTG GTGTTCTCG	960
ACACGATCAA TAGGGCAAAAC CCGGTGCCGG GGAGAGGCCG CATCGAGGCG ACCAACCCGT	1020
GCGGGGAGGT CCCACTGCTG CCTTACGAGT CATGTAATCT CGGCTCGATC AACCTGCC	1080
GGATGCTCGC CGACGGTCGC GTGACTGGG ACCGGCTCGA GGAGGTCGCC GGTGTGGCGG	1140
TGCGGTTCT TGATGACGTC ATCGATGTCA GCCGCTACCC CTTCCCCGAA CTGGGTGAGG	1200
CGGCCCCGCGC CACCCGCAAG ATCGGGCTGG GAGTCATGGG TTTGGCGGAA CTGCTTGCG	1260
CACTGGGTAT TCCGTACGAC AGTGAAGAAG CCGTGCCTT AGCCACCCGG CTCATGCGTC	1320
GCATACAGCA GGCAGCGCAC ACGGCATCGC GGAGGCTGGC CGAAGAGCGG GGCGCATTC	1380
CGGCCTTCAC CGATAGCCGG TTCGCGCGGT CGGGCCCGAG GCGCAACGCA CAGGTACCT	1440
CCGTCGCTCC GACGGGCA	1458

## (2) INFORMATION FOR SEQ ID NO:9:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 862 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

ACGGTGTAA	60
CGTGCTGGAT	
CTGGAACCGC	
GTGGCCGCT	
ACCTACCGAG	
ATCTACTGGC	
GGCGCAGGGG	120
GCTGGCCCTG	
GGCATCGCGG	
TGTCGTAGT	
CGGGATCGCG	
GTGGCCATCG	
TCATCGCCTT	
CGTCGACAGC	180
AGCGCCGGTG	
CCAAACCGGT	
CAGCGCCGAC	
AAGCCGGCCT	
CCGCCAGAG	240
CCATCCGGGC	
TCGCCGGCAC	
CCCAAGCACC	
CCAGCCGGCC	
GGGCAAACCG	
AAGGTAACGC	300
CGCCGCGGCC	
CCGCCGCAGG	
GCCAAAACCC	
CGAGACACCC	
ACGCCAACCG	
CCGCGGTGCA	360
GCCGCCGCCG	
GTGCTCAAGG	
AAGGGACGA	
TTGCCCGAT	
TCGACGCTGG	
CCGTCAAAGG	420
TTTGACCAAC	
GCGCCGCAGT	
ACTACGTCGG	
CGACCAAGCCG	
AAGTTCACCA	
TGGTGGTCAC	480
CAACATCGGC	
CTGGTGTCC	
GTAAACGCGA	
CGTTGGGGCC	
GCGGTGTTGG	
CCGCCTACGT	540
TTACTCGCTG	
GACAACAAGC	
GGTTGTGGTC	
CAACCTGGAC	
TGCGCGCCCT	
CGAATGAGAC	600
GCTGGTCAAG	
ACGTTTTCCC	
CCGGTGAGCA	
GGTAACGACC	
GCGGTGACCT	
GGACCGGGAT	660
GGGATCGGCG	
CCGCGCTGCC	
CATTGCCGCG	
GCCGGCGATC	
GGGCGGGCA	
CCTACAATCT	720
CGTGGTACAA	
CTGGGCAATC	
TGCGCTCGCT	
GCCGGTTCCG	
TTCATCCTGA	
ATCAGCCGCC	780
GCCGCCGCC	
GGGCCGGTAC	
CCGCTCCGGG	
TCCAGCGCAG	
GCGCCTCCGC	
CGGAGTCTCC	840
CGCGCAAGGC	
GGATAATTAT	
TGATCGCTGA	
TGGTCGATT	
CGCCAGCTGT	
GACAACCCCT	862
CGCCTCGTGC	
CG	

## (2) INFORMATION FOR SEQ ID NO:10:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 622 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

TTGATCAGCA CCGGCAAGGC GTCACATGCC TCCCTGGGTG TGCAGGTGAC CAATGACAAA	60
GACACCCCGG GCGCCAAGAT CGTCGAAGTA GTGGCCGGTG GTGCTGCCGC GAACGCTGGA	120
GTGCCGAAGG GCGTCGTTGT CACCAAGGTC GACGACCGCC CGATCAACAG CGCGGACGCG	180
TTGGTTGCCG CCGTGC GGTC CAAAGCGCCG GGCGCCACGG TGGCGCTAAC CTTTCAGGAT	240
CCCTCGGGCG GTAGCCGCAC AGTGCAAGTC ACCCTCGGCA AGGCGGAGCA GTGATGAAGG	300
TCGCCGCGCA GTGTTCAAAG CTCGGATATA CGGTGGCACC CATGGAACAG CGTGC GGAGT	360
TGGTGGTTGG CCGGGCACTT GTCGTGTCG TTGACGATCG CACGGCGCAC GGCGATGAAG	420
ACCACAGCGG GCCGCTTGTC ACCGAGCTGC TCACCGAGGC CGGGTTTGTT GTCGACGGCG	480
TGGTGGCGGT GTCGGCCGAC GAGGTCGAGA TCCGAAATGC GCTGAACACA GCGGTGATCG	540
GCGGGGTGGA CCTGGTGGTG TCGGTCGGCG GGACCGGNGT GACGNCTCGC GATGTCACCC	600
CGGAAGCCAC CCGNGACATT CT	622

## (2) INFORMATION FOR SEQ ID NO:11:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1200 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GGCGCAGCGG TAAGCCTGTT GGCCGCCGGC AACTGGTGT TGACAGCATG CGGCGGTGGC	60
ACCAACAGCT CGTCGTCA GG CGCAGGCGGA ACGTCTGGGT CGGTGCACTG CGGCGGCAAG	120
AAGGAGCTCC ACTCCAGCGG CTCGACCGCA CAAGAAAATG CCATGGAGCA GTTCGTCTAT	180

GCCTACGTGC GATCGTGCC	240
GGGCTACACG TTGGACTACA	
ACGCCAACGG GTCCGGTGCC	
GGGGTGACCC AGTTTCTCAA	300
CAACGAAACC GATTCGCCG	
GCTGGATGT CCCGTTGAAT	
CCGTCGACCG GTCAACCTGA	360
CCGGTCGGCG GAGCGGTGCG	
GTTCCCCGGC ATGGGACCTG	
CCGACGGTGT TCGGCCGAT	420
CGCGATCACCA TACAATATCA	
AGGGCGTGAG CACGCTGAAT	
CTTGACGGAC CCACTACCGC	480
CAAGATTTC AACGGCACCA	
TCACCGTGTG GAATGATCCA	
CAGATCCAAG CCCTCAACTC	540
CGGCACCGAC CTGCCGCCAA	
CACCGATTAG CGTTATCTTC	
CGCAGCGACA AGTCCGGTAC	600
GTCGGACAAAC TTCCAGAAAT	
ACCTCGACGG TGTATCCAAC	
GGGGCGTGGG GCAAAGGCGC	660
CAGCGAAACG TTCAGCGGGG	
GCGTCGGCGT CGGCGCCAGC	
GGGAACAAACG GAACGTCGGC	720
CCTACTGCAG ACGACCGACG	
GGTCGATCAC CTACAACGAG	
TGGTCGTTTG CGGTGGTAA	780
GCAGTTGAAC ATGGCCAGA	
TCATCACGTC GGCGGGTCCG	
GATCCAGTGG CGATCACCAC	840
CGAGTCGGTC GGTAAGACAA	
TCGCCCCGGC CAAGATCATG	
GGACAAGGCA ACGACCTGGT	900
ATTGGACACG TCGTCGTTCT	
ACAGACCCAC CCAGCCTGGC	
TCTTACCGA TCGTGCTGGC	960
GACCTATGAG ATCGTCTGCT	
CGAAATACCC GGATGCGACG	
ACCGGTACTG CGGTAAAGGGC	1020
GTTTATGCAA GCCGCGATTG	
GTCCAGGCCA AGAAGGCCTG	
GACCAATACG GCTCCATTCC	1080
GTTGCCAAA TCGTTCCAAG	
CAAATTGGC GGCGCGGGTG	
AATGCTATTT CTTGACCTAG	1140
TGAAGGGAAT TCGACGGTGA	
GCGATGCCGT TCCGCAGGTA	
GGGTCGCAAT TTGGGCCGTA	1200
TCAGCTATTG CGGCTGCTGG	
GCCGAGGCAG GATGGGCGAG	

## (2) INFORMATION FOR SEQ ID NO:12:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1155 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GCAAGCAGCT	GCAGGGTCGTG	CTGTTCGACG	AACTGGGCAT	GCCGAAGACC	AAACGCCA	60
AGACCGGCTA	CACCACGGAT	GCCGACGCGC	TGCAGTCGTT	GTTCGACAAG	ACCGGGCATC	120
CGTTTCTGCA	ACATCTGCTC	GCCCACCGCG	ACGTCACCCG	GCTCAAGGTC	ACCGTCGACG	180
GGTTGCTCCA	AGCGGTTGCC	GCCGACGGCC	GCATCCACAC	CACGTTAAC	CAGACGATCG	240
CCGCGACCGG	CCGGCTCTCC	TCGACCGAAC	CCAACCTGCA	GAACATCCCG	ATCCGCACCG	300
ACGCGGGCCG	GCGGATCCGG	GACGCGTTCG	TGGTCGGGGA	CGGTTACGCC	GAGTTGATGA	360
CGGCCGACTA	CAGCCAGATC	GAGATGCGGA	TCATGGGGCA	CCTGTCCGGG	GACGAGGGCC	420
TCATCGAGGC	GTTCAACACC	GGGGAGGACC	TGTATTGTT	CGTCGCGTCC	CGGGTGTTG	480
GTGTGCCCAT	CGACGAGGTC	ACCGGGAGT	TGCGCGCCG	GGTCAAGGCG	ATGCTCTACG	540
GGCTGGTTA	CGGGTTGAGC	GCCTACGGCC	TGTCGAGCA	GTTGAAAATC	TCCACCGAGG	600
AAGCCAACGA	GCAGATGGAC	GCGTATTCG	CCCGATTGCG	CGGGGTGCGC	GACTACCTGC	660
GCGCCGTAGT	CGAGCGGGCC	CGCAAGGACG	GCTACACCTC	GACGGTGCTG	GGCCGTCGCC	720
GCTACCTGCC	CGAGCTGGAC	AGCAGCAACC	GTCAAGTGC	GGAGGCCGCC	GAGCGGGCGG	780
CGCTGAACGC	GCCGATCCAG	GGCAGCGCGG	CCGACATCAT	CAAGGTGGCC	ATGATCCAGG	840
TCGACAAGGC	GCTCAACGAG	GCACAGCTGG	CGTCGCGCAT	GCTGCTGCAG	GTCCACGACG	900
AGCTGCTGTT	CGAAATCGCC	CCCGGTGAAC	GCGAGCGGGT	CGAGGCCCTG	GTGCGCGACA	960
AGATGGCGG	CGCTTACCCG	CTCGACGTCC	CGCTGGAGGT	GTCGGTGGGC	TACGGCCGCA	1020
GCTGGGACGC	GGCGGCGCAC	TGAGTGCCGA	GCGTGCATCT	GGGGCGGGAA	TTCGGCGATT	1080
TTTCCGCCCT	GAGTTCACGC	TCGGCGCAAT	CGGGACCGAG	TTTGTCCAGC	GTGTACCCGT	1140
CGAGTAGCCT	CGTCA					1155

## (2) INFORMATION FOR SEQ ID NO:13:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1771 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GAGCGCCGTC	TGGTGTGTTGA	ACGGTTTAC	CGGTCGGCAT	CGGCACGGGC	GTTGCCGGT	60
TCGGGCCTCG	GGTTGGCGAT	CGTCAAACAG	GTGGTGCTCA	ACCACGGCGG	ATTGCTGCGC	120
ATCGAAGACA	CCGACCCAGG	CGGCCAGCCC	CCTGGAACGT	CGATTTACGT	GCTGCTCCCC	180
GGCCGTCGGA	TGCCGATTCC	GCAGCTTCCC	GGTGCACGG	CTGGCGCTCG	GAGCACGGAC	240
ATCGAGAACT	CTCGGGGTTC	GGCGAACGTT	ATCTCAGTGG	AATCTCAGTC	CACGCGCGCA	300
ACCTAGTTGT	GCAGTTACTG	TTGAAAGCCA	CACCCATGCC	AGTCCACGCA	TGGCCAAGTT	360
GGCCCGAGTA	GTGGGCCTAG	TACAGGAAGA	GCAACCTAGC	GACATGACGA	ATCACCCACG	420
GTATTGCCA	CCGCCGCAGC	AGCCGGGAAC	CCCAGGTTAT	GCTCAGGGGC	AGCAGCAAAC	480
GTACAGCCAG	CAGTCGACT	GGCGTTACCC	ACCGTCCCCG	CCCCCGCAGC	CAACCCAGTA	540
CCGTCAACCC	TACGAGGCCT	TGGGTGGTAC	CCGGCCGGGT	CTGATACTG	GCGTGATTCC	600
GACCATGACG	CCCCCTCCTG	GGATGGTCG	CCAACGCCCT	CGTGCAGGCA	TGTTGGCCAT	660
CGGCGCGGTG	ACGATAGCGG	TGGTGTCCGC	CGGCATCGGC	GGCGCGGCCG	CATCCCTGGT	720
CGGGTTCAAC	CGGGCACCCG	CGGGCCCCAG	CGGCAGGCCA	GTGGCTGCCA	GCGCGGCC	780
AAGCATCCCC	GCAGCAAACA	TGCCGCCGGG	GTCGGTCGAA	CAGGTGGCGG	CCAAGGTGGT	840
GCCCAGTGTC	GTCATGTTGG	AAACCGATCT	GGGCCGCCAG	TGGGAGGAGG	GCTCCGGCAT	900
CATTCTGTCT	GCCGAGGGGC	TGATCTTGAC	CAACAACCAC	GTGATCGCGG	CGGCCGCCAA	960

GCCTCCCTG GGCAGTCCGC CGCCGAAAAC GACGGTAACC TTCTCTGACG GGCGGACCGC	1020
ACCCCTTCACG GTGGTGGGGG CTGACCCCAC CAGTGATATC GCCGTCGTCC GTGTTCAGGG	1080
CGTCTCCGGG CTCACCCCGA TCTCCCTGGG TTCCTCTCG GACCTGAGGG TCGGTCAAGCC	1140
GGTGCTGGCG ATCGGGTCGC CGCTCGGTTT GGAGGGCACCC GTGACCACGG GGATCGTCAG	1200
CGCTCTCAAC CGTCCAGTGT CGACGACCGG CGAGGCCGGC AACCAGAACAA CCGTGCTGGA	1260
CGCCATTCAAG ACCGACGCCG CGATCAACCC CGGTAACCTCC GGGGGCGCGC TGGTGAACAT	1320
GAACGCTCAA CTCGTCGGAG TCAACTCGGC CATTGCCACG CTGGGCGCGG ACTCAGCCGA	1380
TGCGCAGAGC GGCTCGATCG GTCTCGGTTT TGCGATTCCA GTCGACCAGG CCAAGCGCAT	1440
CGCCGACGAG TTGATCAGCA CCGGCAAGGC GTCACATGCC TCCCTGGGTG TGCAAGGTGAC	1500
CAATGACAAA GACACCCCGG GCGCCAAGAT CGTCGAAGTA GTGGCCGGTG GTGCTGCCGC	1560
GAACGCTGGA GTGCCGAAGG GCGTCGTTGT CACCAAGGTC GACGACCGCC CGATCAACAG	1620
CGCGGACGCG TTGGTTGCCG CCGTGCAGTC CAAAGCGCCG GGCGCCACGG TGGCGCTAAC	1680
CTTCAGGAT CCCTCGGGCG GTAGCCGCAC AGTGAAGTC ACCCTCGGCA AGGCGGAGCA	1740
GTGATGAAGG TCGCCGCGCA GTGTTCAAAG C	1771

## (2) INFORMATION FOR SEQ ID NO:14:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1058 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CTCCACCGCG GTGGCGGCCG CTCTAGAACT AGTGGATCCC CCGGGCTGCA GGAATTCCGC	60
ACGAGGATCC GACGTCGCAG GTTGTGAAAC CCGCCGCCGC GGAAGTATCG GTCCATGCCT	120

AGCCC GGCGA CGGCGAGCGC CGGAATGGCG CGAGTGAGGA GGC GGGCAAT TTGGC GGGGC	180
CCGGCGACGG CGAGCGCCGG AATGGCGCGA GTGAGGAGGC GGGCAGTCAT GCCCAGCGTG	240
ATCCAATCAA CCTGCATTG GCCTGC GGGC CCATTGACA ATCGAGGTAG TGAGCGCAA	300
TGAATGATGG AAAAC GGGCG GTGACGTCCG CTGTTCTGGT GGTGCTAGGT GCCTGCCTGG	360
CGTTGTGGCT ATCAGGATGT TCTTCGCCGA AACCTGATGC CGAGGAACAG GGTGTTCCCG	420
TGAGCCC GAC GGC GTCC GAC CCC GCG CTC TCGCC GAGAT CAGG CAGTCG CTTGATGCGA	480
CAA AAGGGTT GACCAGCGTG CACGTAGCGG TCCGAACA AC CGGGAAAGTC GACAGCTTGC	540
TGGGTATTAC CAGTGCCGAT GTCGACGTCC GGGCAATCC GCTCGCGCA AAGGGCGTAT	600
GCACCTACAA CGACGAGCAG GGTGTCCCGT TTCGGGTACA AGGC GACA AC ATCTCGGTGA	660
AACTGTTCGA CGACTGGAGC AATCTCGGCT CGATTCTGA ACTGTCAACT TCACGCGTGC	720
TCGATCCTGC CGCTGGGTG ACGCAGCTGC TGTCCGGTGT CACGAACCTC CAAGCGCAAG	780
GTACCGAAGT GATA GACGGA ATTTCGACCA CCAA AATCAC CGGGACC ATC CCC GCGAGCT	840
CTGTCAAGAT GCTTGATCCT GGCGCCAAGA GTGCAAGGCC GGCGACCGTG TGGATTGCC	900
AGGACGGCTC GCACCACCTC GTCCGAGCGA GCATCGACCT CGGATCCGGG TCGATT CAGC	960
TCACGCGAGTC GAAATGGAAC GAACCCGTCA ACGTCGACTA GGCGGAAGTT GCGTCGACGC	1020
GTTGNTCGAA ACGCCCTTGT GAACGGTGTC AACGGNAC	1058

## (2) INFORMATION FOR SEQ ID NO:15:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 542 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GAATTCGGCA CGAGAGGTGA TCGACATCAT CGGGACCAGC CCCACATCCT GGGAACAGGC	60
GGCGGCGGAG GCGGTCCAGC GGGCGCGGGA TAGCGTCGAT GACATCCGCG TCGCTCGGGT	120
CATTGAGCAG GACATGGCCG TGGACAGCGC CGGCAAGATC ACCTACCGCA TCAAGCTCGA	180
AGTGTGTTTC AAGATGAGGC CGGCGCAACC GCGCTAGCAC GGGCCGGCGA GCAAGACGCA	240
AAATCGCACG GTTGCGGTT GATTGCGCG ATTTTGTGTC TGCTCGCCGA GGCTTACCAAG	300
GCGCGGCCCA GGTCCGCGTG CTGCCGTATC CAGGC GTGCA TCGCGATTCC GGCGGCCACG	360
CCGGAGTTAA TGCTTCGCGT CGACCCGAAC TGGGCGATCC GCCGGNGAGC TGATCGATGA	420
CCGTGGCCAG CCCGTCGATG CCCGAGTTGC CCGAGGAAAC GTGCTGCCAG GCCGGTAGGA	480
AGCGTCCGTA GGC GGCGGTG CTGACCGGCT CTGCCCTGCCGC CCTCAGTGCG GCCAGCGAGC	540
GG	542

## (2) INFORMATION FOR SEQ ID NO:16:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 913 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CGGTGCCGCC CGCGCCTCCG TTGCCCTCAT TGCCGCCGTC GCCGATCAGC TGCGCATCGC	60
CACCATCACC GCCTTGCGC CGGGCACCGC CGGTGGCGCC GGGGCCGCCG ATGCCACCGC	120
TTGACCCCTGG CGGCCGGCGC CGCCATTGCC ATACAGCACC CCGCCGGGGG CACCGTTACC	180
GCGTCGCCA CGTCGCCGC CGCTGCCGTT TCAGGCCGGG GAGGCCGAAT GAACCGCCGC	240
CAAGCCCGCC GCCGGCACCG TTGCCGCCTT TTCCGCCCGC CCCGCCGGCG CCGCCAATTG	300

CCGAACAGCC AMGCACCGTT GCCGCCAGCC CCGCCGCCGT TAACGGCGCT GCCGGGCC	360
GCCGCCGGAC CGGCCATTAC CGCCGTTCCC GTTCGGTGCC CGGCCGTTAC CGGCCGCC	420
GTTTGCAGCC AATATTCGGC GGGCACCGCC AGACCCGCCG GGGCCACCAT TGCCGCCGG	480
CACCGAAACA ACAGCCAAC GGTGCCGCCG GCCCCGCCGT TTGCCGCCAT CACCGGCCAT	540
TCACCGCCAG CACCGCCGTT AATGTTTATG AACCCGGTAC CGCCAGCGCG GCCCCTATTG	600
CCGGGCGCCG GAGNGCGTGC CCGCCGGCGC CGCCAACGCC CAAAAGCCCG GGGTTGCCAC	660
CGGCCCCGCC GGACCCACCG GTCCCCCGA TCCCCCGTT GCCGCCGGTG CCGCCGCCAT	720
TGGTGCTGCT GAAGCCGTTA GCGCCGGTTC CGCSGGTTCC GGCGGTGGCG CCNTGGCCGC	780
CGGCCCCGCC GTTGCCGTAC AGCCACCCCC CGGTGGCGCC GTTGCCGCCA TTGCCGCCAT	840
TGCCGCCGTT GCCGCCATTG CCGCCGTTCC CGCCGCCACC GCCGGNTTGG CCGCCGGCGC	900
CGCCGGCGGC CGC	913

## (2) INFORMATION FOR SEQ ID NO:17:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1872 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GACTACGTTG GTGTAGAAAA ATCCTGCCGC CGGGACCCCTT AAGGCTGGGA CAATTCTGA	60
TAGCTACCCC GACACAGGAG GTTACGGGAT GAGCAATTG CGCCGCCGCT CACTCAGGTG	120
GTCATGGTTG CTGAGCGTGC TGGCTGCCGT CGGGCTGGGC CTGGCCACGG CGCCGGCCCA	180
GGCGGGCCCCG CGGGCCTTGT CGCAGGACCG GTTCGCCGAC TTCCCCGCGC TGCCCCCTCGA	240

CCCGTCCGCG ATGGTCGCCA AAGTGGGCC ACAGGTGGTC AACATCAACA CCAAACCTGGG	300
CTACAACAAC GCCGTGGCG CGGGGACCGG CATCGTCATC GATCCCACG GTGTCGTGCT	360
GACCAACAAC CACGTGATCG CGGGCGCCAC CGACATCAAT GCGTTCAGCG TCGGCTCCGG	420
CCAAACCTAC GGCGTCGATG TGGTGGGTA TGACCGCACC CAGGATGTGCG CGGTGCTGCA	480
GCTGCGCGGT GCCGGTGGCC TGCGTCGGC GGCGATCGGT GGCGGCGTCG CGGTTGGTGA	540
GCCCCGTCGTC GCGATGGGCA ACAGCGGTGG GCAGGGCGGA ACGCCCCGTG CGGTGCCTGG	600
CAGGGTGGTC GCGCTGGCC AAACCGTGCA GGCGTCGGAT TCGCTGACCG GTGCCGAAGA	660
GACATTGAAC GGGTTGATCC AGTTCGATGC CGCAATCCAG CCCGGTGATT CGGGCGGGCC	720
CGTCGTCAAC GGCCTAGGAC AGGTGGTCGG TATGAACACG GCCGCGTCCG ATAACTTCCA	780
GCTGTCCAG GGTGGGCAGG GATTGCCAT TCCGATCGGG CAGGCGATGG CGATCGCGGG	840
CCAAATCCGA TCGGGTGGGG GGTCAACCCAC CGTTCATATC GGGCCTACCG CCTTCCTCGG	900
CTTGGGTGTT GTCGACAACA ACGGCAACGG CGCACGAGTC CAACGCGTGG TCGGAAGCGC	960
TCCGGCGGCA AGTCTCGGCA TCTCCACCGG CGACGTGATC ACCGCGGTG ACGGCGCTCC	1020
GATCAACTCG GCCACCGCGA TGGCGGACGC GCTTAACGGG CATCATCCCG GTGACGTAT	1080
CTCGGTGAAC TGGCAAACCA AGTCGGCGG CACGCGTACA GGGAACGTGA CATTGGCCGA	1140
GGGACCCCCG GCCTGATTTG TCGCGGATAAC CACCCGCCGG CGGGCCAATT GGATTGGCGC	1200
CAGCCGTGAT TGCCGCGTGA GCCCCCCAGT TCCGTCTCCC GTGCGCGTGG CATTGTGGAA	1260
GCAATGAACG AGGCAGAACAA CAGCGTTGAG CACCCCTCCCG TGCAAGGGCAG TTACGTCGAA	1320
GGCGGTGTGG TCGAGCATCC GGATGCCAAG GACTTCGGCA GCGCCGCCGC CCTGCCCGCC	1380
GATCCGACCT GGTTTAAGCA CGCCGTCTTC TACGAGGTGC TGGTCCGGGC GTTCTTCGAC	1440
GCCAGCGCGG ACGGTTCCGN CGATCTCGGT GGACTCATCG ATCGCCTCGA CTACCTGCAG	1500
TGGCTTGGCA TCGACTGCAT CTGTTGCCGC CGTTCTACG ACTCACCGCT GCGCGACGGC	1560
GGTTACGACA TTCGCGACTT CTACAAGGTG CTGCCCCAAT TCGGCACCGT CGACGATTTC	1620

GTCGCCCTGG TCGACACCGC TCACCGGCGA GGTATCCGCA TCATCACCGA CCTGGTGATG	1680
AATCACACCT CGGAGTCGCA CCCCTGGTTT CAGGAGTCCC GCCGCGACCC AGACGGACCG	1740
TACGGTGACT ATTACGTGTG GAGCGACACC AGCGAGCGCT ACACCGACGC CCGGATCATC	1800
TTCGTCGACA CGGAAGAGTC GAACTGGTCA TTGATCCTG TCCGCCGACA GTTNCTACTG	1860
GCACCGATTCTT	1872

## (2) INFORMATION FOR SEQ ID NO:18:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1482 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CTTCGCGCAA ACCTGATGCC GAGGAACAGG GTGTTCCCGT GAGCCCGACG GCGTCCGACC	60
CCCGCGCTCCT CGCCGAGATC AGGCAGTCGC TTGATGCGAC AAAAGGGTTG ACCAGCGTGC	120
ACGTAGCGGT CCGAACAAACC GGGAAAGTCG ACAGCTTGCT GGGTATTACC AGTGCCGATG	180
TCGACGTCCG GGCCAATCCG CTCGCGGCAA AGGGCGTATG CACCTACAAC GACGAGCAGG	240
GTGTCCCGTT TCGGGTACAA GGCGACAACA TCTCGGTGAA ACTGTTGAC GACTGGAGCA	300
ATCTCGGCTC GATTTCTGAA CTGTCAACTT CACGCGTGCT CGATCCTGCC GCTGGGGTGA	360
CGCAGCTGCT GTCCGGTGTC ACGAACCTCC AAGCGCAAGG TACCGAAGTG ATAGACGGAA	420
TTTCGACCAAC CAAAATCACC GGGACCATCC CCGCGAGCTC TGTCAAGATG CTTGATCCTG	480
GCGCCAAGAG TGCAAGGCCG GCGACCGTGT GGATTGCCA GGACGGCTCG CACCACCTCG	540
TCCGAGCGAG CATCGACCTC GGATCCGGGT CGATTCAAGCT CACGCGATCG AAATGGAACG	600

AACCCGTCAA CGTCGACTAG GCCGAAGTTG CGTCGACGCG TTGCTCGAAA CGCCCTTGTG	660
AACGGTGTCA ACGGCACCCG AAAACTGACC CCCTGACGGC ATCTGAAAAT TGACCCCTA	720
GACCGGGCGG TTGGTGGTTA TTCTTCGGTG GTTCCGGCTG GTGGGACGCG GCCGAGGTG	780
CGGTCTTGA GCCGGTAGCT GTCGCCTTG AGGGCGACGA CTTCAGCATG GTGGACGAGG	840
CGGTCGATCA TGGCGGCAGC AACGACGTCG TCGCCGCCGA AAACCTCGCC CCACCGGCCG	900
AAGGCCTTAT TGGACGTGAC GATCAAGCTG GCCCCGCTCAT ACCGGGAGGA CACCAGCTGG	960
AAGAAGAGGT TGGCGGCCTC GGGCTCAAAC GGAATGTAAC CGACTTCGTC AACCAACCAGG	1020
AGCGGATAGC GGCCAAACCG GGTGAGTTCG GCGTAGATGC GCCCGGCGTG GTGAGCCTCG	1080
GCGAACCGTG CTACCCATTG GGCGGCGGTG GCGAACAGCA CCCGATGACC GGCCTGACAC	1140
GCGCGTATCG CCAGGCCGAC CGCAAGATGA GTCTTCCCGG TGCCAGGCGG GGCCCAAAAA	1200
CACGACGTTA TCGCGGGCGG TGATGAAATC CAGGGTGCCC AGATGTGCGA TGGTGTGCG	1260
TTTGAGGCCA CGAGCATGCT CAAAGTCGAA CTCTTCCAAC GACTTCCGAA CGGGGAAGCG	1320
GGCGGCGCGG ATGCGGCCCT CACCACCATG GGAATCCCGG GCTGACACTT CCCGCTGCAG	1380
GCAGGCGGCC AGGTATTCTT CGTGGCTCCA GTTCTCGCG CGGGCGCGAT CGGCCAGCCG	1440
GGACACTGAC TCACGCAGGG TGGGAGCTTT CAATGCTTT GT	1482

## (2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 876 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GAATTCGGCA CGAGCCGGCG ATAGCTTCTG GGCGCGGCC GACCAAGATGG CTCGAGGGTT	60
---	----

CGTGCTCGGG	GCCACCGCCG	GGCGCACCAAC	CCTGACCGGT	GAGGGCCTGC	AACACGCCGA	120	
CGGTCACTCG	T	TGCTGCTGG	ACGCCACCAA	CCC GGCGGTG	GTTGCCTACG	ACCCGGCCTT	180
CGCCTACGAA	ATCGGCTACA	TCGNNGAAAG	CGGACTGGCC	AGGATGTGCG	GGGAGAACCC	240	
GGAGAACATC	TTCTTCTACA	TCACCGTCTA	CAACGAGCCG	TACGTGCAGC	CGCCGGAGCC	300	
GGAGAACATTG	GATCCCGAGG	GCGTGCTGGG	GGGTATCTAC	CGNTATCACG	CGGCCACCAGA	360	
GCAACGCACC	AACAAGGNGC	AGATCCTGGC	CTCCGGGTA	GCGATGCCCG	CGGCGCTGCG	420	
GGCAGCACAG	ATGCTGGCCG	CCGAGTGGGA	TGTCGCCGCC	GACGTGTGGT	CGGTGACCAG	480	
TTGGGGCGAG	CTAAACCGCG	ACGGGGTGGT	CATCGAGACC	GAGAAGCTCC	GCCACCCCGA	540	
TCGGCCGGCG	GGCGTGCCCT	ACGTGACGAG	AGCGCTGGAG	AATGCTCGGG	GCCCGGTGAT	600	
CGCGGTGTCG	GACTGGATGC	GCGCGGTCCC	CGAGCAGATC	CGACCGTGGG	TGCCGGGCAC	660	
ATACCTCACG	TTGGGCACCG	ACGGGTTCTGG	TTTTCCGAC	ACTCGGCCCG	CCGGTCGTCG	720	
TTACTTCAAC	ACCGACGCCG	AATCCCAGGT	TGGTCGCGGT	TTTGGGAGGG	GTTGGCCGGG	780	
TCGACGGGTG	AATATCGACC	CATTGGTGC	CGGTCGTGGG	CCGCCCCGCC	AGTTACCCGG	840	
ATTCGACGAA	GGTGGGGGGT	TGCGCCCGAN	TAAGTT			876	

## (2) INFORMATION FOR SEQ ID NO:20:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1021 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

ATCCCCCCGG	GCTGCAGGAA	TTCGGCACGA	GAGACAAAAT	TCCACGCGTT	AATGCAGGAA	60
------------	------------	------------	------------	------------	------------	----

CAGATTATA	ACGAATTAC	AGCGGCACAA	CAATATGTCG	CGATCGCGGT	TTATTCGAC	120
AGCGAAGACC	TGCCGCAGTT	GGCGAACAT	TTTACAGCC	AAGCGGTGA	GGAACGAAAC	180
CATGCAATGA	TGCTCGTGCA	ACACCTGCTC	GACCGCGACC	TTCGTGTGA	AATTCCCGC	240
GTAGACACGG	TGCGAAACCA	GTTCGACAGA	CCCCGCGAGG	CACTGGCGCT	GGCGCTCGAT	300
CAGGAACGCA	CAGTCACCGA	CCAGGTCGGT	CGGCTGACAG	CGGTGGCCCG	CGACGAGGGC	360
GATTTCTCG	GCGAGCAGTT	CATGCAGTGG	TTCTTGCAGG	AACAGATCGA	AGAGGTGGCC	420
TTGATGGCAA	CCCTGGTGCG	GGTTGCCGAT	CGGGCCGGGG	CCAACCTGTT	CGAGCTAGAG	480
AACTTCGTCG	CACGTGAAGT	GGATGTGGCG	CCGGCCGCAT	CAGGCGCCCC	GCACGCTGCC	540
GGGGGCCGCC	TCTAGATCCC	TGGGGGGGAT	CAGCGAGTGG	TCCCGTTCGC	CCGCCCCGTCT	600
TCCAGGCCAGG	CCTTGGTGCG	GCCGGGGTGG	TGAGTACCAA	TCCAGGCCAC	CCCGACCTCC	660
CGGNAAAAGT	CGATGTCTC	GTACTCATCG	ACGTTCCAGG	AGTACACCGC	CCGGCCCTGA	720
GCTGCCGAGC	GGTCAACGAG	TTGCGGATAT	TCCTTAACG	CAGGCAGTGA	GGGTCCCACG	780
CGGGTTGGCC	CGACCGCCGT	GGCCGCACTG	CTGGTCAGGT	ATCGGGGGGT	CTTGGCGAGC	840
AACAACGTCG	GCAGGAGGGG	TGGAGCCCGC	CGGATCCGCA	GACCGGGGGG	GCGAAAACGA	900
CATCAACACC	GCACGGGATC	GATCTGCGGA	GGGGGGTGCG	GGAATACCGA	ACCGGTGTAG	960
GAGCGCCAGC	AGTTGTTTTT	CCACCAGCGA	AGCGTTTCG	GGTCATCGGN	GGCNNTTAAG	1020
T						1021

## (2) INFORMATION FOR SEQ ID NO:21:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 321 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CGTGCCGACG AACGGAAGAA CACAACCATG AAGATGGTGA AATCGATCGC CGCAGGTCTG	60
ACCGCCGCGG CTGCAATCGG CGCCGCTGCG GCCGGTGTGA CTTCGATCAT GGCTGGCGGN	120
CCGGTCGTAT ACCAGATGCA GCCGGTCGTC TTCCGGCGGC CACTGCCGTT GGACCCGGNA	180
TCCGCCCTG ANGTCCCAC CGCCGCCAG TGGACCAGNC TGCTAACAG NCTCGNCGAT	240
CCCAACGTGT CGTTTNGAA CAAGGGNAGT CTGGTCGAGG GNNGNATCGG NGGNANCAG	300
GGNGNGNATC GNCGANCACA A	321

## (2) INFORMATION FOR SEQ ID NO:22:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 373 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

TCTTATCGGT TCCGGTTGGC GACGGGTTTT GGGNGCGGGT GGTAAACCG CTCGGCCAGC	60
CGATCGACGG GCGCGGAGAC GTCGACTCCG ATACTCGCG CGCGCTGGAG CTCCAGGCGC	120
CCTCGGTGGT GNACCGGCAA GGC GTGAAGG AGCCGTTGNA GACCGGGATC AAGGCGATTG	180
ACGCGATGAC CCCGATCGGC CGCGGGCAGC GCCAGCTGAT CATCGGGAC CGCAAGACCG	240
GCAAAAACCG CCGTCTGTGT CGGACACCAT CCTCAAACCA GCGGGAAGAA CTGGGAGTCC	300
GGTGGATCCC AAGAACGAGG TGCGCTTGTG TATACGTTGG CCATCGGGCA AGAAGGGGAA	360
CTTACCATCG CCG	373

## (2) INFORMATION FOR SEQ ID NO:23:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 352 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GTGACGCCGT GATGGGATT C TGGGGCGGGG CCGGTCGCT GGCGGTGGTG GATCAGCAAC	60
TGGTTACCCG GGTGCCGCAA GGCTGGTCGT TTGCTCAGGC AGCCGCTGTG CCGGTGGTGT	120
TCTTGACGGC CTGGTACGGG TTGGCCGATT TAGCCGAGAT CAAGGCAGGC GAATCGGTGC	180
TGATCCATGC CGGTACCGGC GGTGTGGGCA TGGCGGCTGT GCAGCTGGCT CGCCAGTGGG	240
GCGTGGAGGT TTTCGTCACC GCCAGCCGTG GNAAGTGGGA CACCGCTGGC GCCATNGNGT	300
TTGACGACGA NCCATATCGG NGATTCCCNC ACATNCGAAG TTCCGANGGA GA	352

## (2) INFORMATION FOR SEQ ID NO:24:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 726 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GAAATCCGCG TTCATTCCGT TCGACCAGCG GCTGGCGATA ATCGACGAAG TGATCAAGCC	60
GCGGTTCGCG GCGCTCATGG GTCACAGCGA GTAATCAGCA AGTTCTCTGG TATATCGCAC	120
CTAGCGTCCA GTTGCTTGCC AGATCGCTTT CGTACCGTCA TCGCATGTAC CGGTTCGCGT	180
GCCGCACGCT CATGCTGGCG GCGTGCATCC TGGCCACGGG TGTGGCGGGT CTCGGGGTCG	240

GCGCGCAGTC CGCAGCCAA ACCGCGCCGG TGCCCAGTA CTACTGGTGC CCGGGGCAGC	300
CTTTCGACCC CGCATGGGG CCCAACTGGG ATCCCTACAC CTGCCATGAC GACTTCCACC	360
GCGACAGCGA CGGCCCCGAC CACAGCCGCG ACTACCCCGG ACCCATCCTC GAAGGTCCCG	420
TGCTTGACGA TCCCGGTGCT GCGCCGCCGC CCCCCGGCTGC CGGTGGCGGC GCATAGCGCT	480
CGTTGACCGG GCCGCATCAG CGAATACGCG TATAAACCGG GGCGTGCCCC CGGCAAGCTA	540
CGACCCCCGG CGGGGCAGAT TTACGCTCCC GTGCCGATGG ATCGCGCCGT CCGATGACAG	600
AAAATAGGCG ACGGTTTG CAACCGCTTG GAGGACGCTT GAAGGGAACC TGTCAATGAAC	660
GGCGACAGCG CCTCCACCAT CGACATCGAC AAGGTTGTTA CCCGCACACC CGTTCGCCGG	720
ATCGTG	726

## (2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 580 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

CGCGACGACG ACGAACGTCG GGCCCACAC CGCCTATGCG TTGATGCAGG CGACCGGGAT	60
GGTCGCCGAC CATATCCAAG CATGCTGGGT GCCCACTGAG CGACCTTTG ACCAGCCGG	120
CTGCCCGATG GCGGCCCGGT GAAGTCATTG CGCCGGGGCT TGTGCACCTG ATGAACCGA	180
ATAGGGAACA ATAGGGGGGT GATTGGCAG TTCAATGTCG GGTATGGCTG GAAATCCAAT	240
GGCGGGGCAT GCTCGCGGCC GACCAGGCTC GCGCAGGCGG GCCAGCCGA ATCTGGAGGG	300
AGCACTCAAT GGCGGCGATG AAGCCCCGGA CCGGCGACGG TCCTTGGAA GCAACTAAGG	360

AGGGGCGCGG CATTGTGATG CGAGTACCAAC TTGAGGGTGG CGGTCGCCTG GTCGTCGAGC	420
TGACACCCGA CGAACGCCGCC GCACTGGGTG ACGAACTCAA AGGCCTTACT AGCTAAGACC	480
AGCCCAACGG CGAATGGTCG GCGTTACCGCG CACACCTTCC GGTAGATGTC CAGTGTCTGC	540
TCGGCGATGT ATGCCAGGA GAACTCTTGG ATACAGCGCT	580

## (2) INFORMATION FOR SEQ ID NO:26:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 160 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

AACGGAGGCG CGGGGGGTTT TGGCGGGGCC GGGGCGGTG GCGGCAACGG CGGGGCCGGC	60
GGTACCGCCG GGTTGTTCGG TGTCGGCGGG GCGGGTGGGG CCGGAGGCAA CGGCATCGCC	120
GGTGTACCGG GTACGTCGGC CAGCACACCG GGTGGATCCG	160

## (2) INFORMATION FOR SEQ ID NO:27:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 272 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GACACCGATA CGATGGTGAT GTACGCCAAC GTTGTGACA CGCTCGAGGC GTTCACGATC	60
CAGCGCACAC CCGACGGCGT GACCATCGGC GATGCGGCC CGTTCGCGGA GGCGGCTGCC	120

AAGGCGATGG GAATCGACAA GCTGCGGGTA ATTCAACCG GAATGGACCC CGTCGTCGCT	180
GAACGCGAAC AGTGGGACGA CGGCAACAAC ACGTTGGCGT TGGCGCCCGG TGTCGTTGTC	240
GCCTACGAGC GCAACGTACA GACCAACGCC CG	272

## (2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 317 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GCAGCCGGTG GTTCTCGGAC TATCTGCGCA CGGTGACGCA GCGCGACGTG CGCGAGCTGA	60
AGCGGATCGA GCAGACGGAT CGCCTGCCGC GGTTCATGCG CTACCTGGCC GCTATCACCG	120
CGCAGGAGCT GAACGTGGCC GAAGCGGCAC GGGTCATCGG GGTGACGCG GGGACGATCC	180
GTTCGGATCT GGCGTGGTTC GAGACGGTCT ATCTGGTACA TCGCCTGCCGC GCCTGGTCGC	240
GGAATCTGAC CGCGAAGATC AAGAAGCGGT CAAAGATCCA CGTCGTCGAC AGTGGCTTCG	300
CGGCCTGGTT GCGCGGG	317

## (2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 182 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GATCGTGGAG CTGTCGATGA ACAGCGTTGC CGGACGCGCG GCGGCCAGCA CGTCGGTGTA	60
GCAGCGCCGG ACCACCTCGC CGGTGGGCAG CATGGTGATG ACCACGTCGG CCTCGGCCAC	120
CGCTTCGGGC GCGCTACGAA ACACCGCGAC ACCGTGCGCG GCGGCCGG AGCGCCCGT	180
GG	182

## (2) INFORMATION FOR SEQ ID NO:30:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 308 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GATCGCGAAG TTTGGTGAGC AGGTGGTCGA CGCGAAAGTC TGGGCGCCTG CGAACGGGT	60
CGGCGTTCAC GAGGCGAAGA CACGCCTGTC CGAGCTGCTG CGGCTCGTCT ACGGCGGGCA	120
GAGGTTGAGA TTGCCCGCCG CGGCGAGCCG GTAGCAAAGC TTGTGCCGCT GCATCCTCAT	180
GAGACTCGGC GGTTAGGCAT TGACCATGGC GTGTACCGCG TGCCCGACGA TTTGGACGCT	240
CCGTTGTCAG ACGACGTGCT CGAACGCTTT CACCGGTGAA GCGCTACCTC ATCGACACCC	300
ACGTTTGG	308

## (2) INFORMATION FOR SEQ ID NO:31:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 267 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

CCGACGACGA GCAACTCACG TGGATGATGG TCGGCAGCGG CATTGAGGAC GGAGAGAAC	60
CGGCCGAAGC TGCCGCGCGG CAAGTGCTCA TAGTGACCGG CCGTAGAGGG CTCCCCGAT	120
GGCACCGGAC TATTCTGGTG TGCCGCTGGC CGGTAAGAGC GGGTAAAAGA ATGTGAGGGG	180
ACACGATGAG CAATCACACC TACCGAGTGA TCGAGATCGT CGGGACCTCG CCCGACGGCG	240
TCGACGCGGC AATCCAGGGC GGTCTGG	267

## (2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 189 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

CTCGTGCCGA AAGAATGTGA GGGGACACGA TGAGCAATCA CACCTACCGA GTGATCGAGA	60
TGTCGGGAC CTCGCCCCAC GGCCTCGACG CGGCAATCCA GGGCGGTCTG GCCCGAGCTG	120
CGCAGACCAT GCGCGCGCTG GACTGGTTCG AAGTACAGTC AATTGAGGC CACCTGGTCG	180
ACGGAGCGG	189

## (2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 851 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

CTGCAGGGTG GCGTGGATGA GCGTCACCGC	60
GGGGCAGGCC GAGCTGACCG CCGCCCAGGT	
CCGGGTTGCT GCGGCGGCCT ACGAGACGGC	120
GTATGGGCTG ACGGTGCCCG CGCCGGTGAT	
CGCCGAGAAC CGTGCTGAAC TGATGATTCT	180
GATAGCGACC AACCTCTTGG GGCAAAACAC	
CCCGGCGATC GCGGTCAACG AGGCCGAATA	240
CGGCGAGATG TGGGCCAAG ACGCCGCCGC	
GATGTTTGGC TACGCCGCGG CGACGGCGAC	300
GGCGACGGCG ACGTTGCTGC CGTTCGAGGA	
GGCGCCGGAG ATGACCAGCG CGGGTGGGCT	360
CCTCGAGCAG GCCGCCGCGG TCGAGGAGGC	
CTCCGACACC GCCGCGGCCGA ACCAGTTGAT	420
GAACAATGTG CCCCAGGCC TGAAACAGTT	
GGCCCAGCCC ACGCAGGGCA CCACGCCCTTC	480
TTCCAAGCTG GGTGGCCTGT GGAAGACGGT	
CTCGCCGCAT CGGTCGCCGA TCAGAACAT	540
GGTGTGATG GCCAACAAACC ACATGTCGAT	
GACCAACTCG GGTGTGTCGA TGACCAACAC	600
CTTGAGCTCG ATGTTGAAGG GCTTGCTCC	
GGCGGGCGGCC GCCCAGGCCG TGCAAACCGC	660
GGCGCAAAAC GGGGTCCGGG CGATGAGCTC	
GCTGGGCAGC TCGCTGGTT CTTCGGGTCT	720
GGGCGGTGGG GTGGCCGCCA ACTTGGGTCG	
GGCGGCCTCG GTACGGTATG GTCACCGGGA	780
TGGCGGAAAA TATGCANAGT CTGGTCGGCG	
GAACGGTGGT CGGGCGTAAG GTTTACCCCC	840
GTTTCTGGA TGCGGTGAAC TTCTGTCAACG	
GAAACAGTTA C	851

## (2) INFORMATION FOR SEQ ID NO:34:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 254 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GATCGATCGG GCGGAAATT GGACCAGATT CGCCTCCGGC GATAACCAA TCAATGAAC	60
CTAGATTAT TCCGTCCAGG GGCCCGAGTA ATGGCTCGCA GGAGAGGAAC CTTACTGCTG	120
CGGGCACCTG TCGTAGGTCC TCGATACGGC GGAAGGCAGTC GACATTTCC ACCGACACCC	180
CCATCCAAAC GTTCGAGGGC CACTCCAGCT TGTGAGCGAG GCGACGCAGT CGCAGGCTGC	240
GCTTGGTCAA GATC	254

## (2) INFORMATION FOR SEQ ID NO:35:

- (1) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 408 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

CGGCACGAGG ATCCTGACCG AAGCGGCCGC CGCCAAGGCG AAGTCGCTGT TGGACCAGGA	60
GGGACGGGAC GATCTGGCGC TGCGGATCGC GGTTCAGCCG GGGGGGTGCG CTGGATTGCG	120
CTATAACCTT TTCTTCGACG ACCGGACGCT GGATGGTGAC CAAACCGCGG AGTCGGTGG	180
TGTCAGGTTG ATCGTGGACC GGATGAGCGC GCCGTATGTG GAAGGCGCGT CGATCGATT	240
CGTCGACACT ATTGAGAAGC AAGGNTTCAC CATCGACAAT CCCAACGCCA CCGGCTCCTG	300
CGCGTGCAGG GATTGTTCA ACTGATAAAA CGCTAGTACG ACCCCGCGGT GCGAACACG	360
TACGAGCACA CCAAGACCTG ACCGCGCTGG AAAAGCAACT GAGCGATG	408

## (2) INFORMATION FOR SEQ ID NO:36:

- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 181 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

GC GG TG TCGG CG GATCCGGC GGGTGGTTGA ACGGCAACGG CG GGGGCCGGC GGGGCCGGCG	60
GG ACCCGGC GC TAACGGTGGT GCCGGCGGC ACG CCTGGTT GTT CGGGGCGC GG CGGGTCCG	120
GC GGN GCGG CACCAATGGT GGN GTCGGCG GGT CCGGCGG ATT TGCTAC GG CAACGGCG	180
G	181

(2) INFORMATION FOR SEQ ID NO:37:

- (i) SEQUENCE CHARACTERISTICS:

  - (A) LENGTH: 290 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

GC GG TG TCGG CG GATCCGGC GGGTGGTTGA ACGGCAACGG CG GTG TCGGC GG CG GGGGCG	60
GCG AC GGC GT CT TGCCGGT GCCGGCGGC AGGGCGGC CT CGT GGGCAG GG CGG CAATG	120
GC GGC GGCTC CAC CGG CGGC AAC CGG CGTC TT GGCGG CGC GG CGG TGGC GG AGG CAACG	180
CCCC CGG AC GG CGG CTT CGGT GG CAAC GG CG GTA AGG GTGG CC AGG GCG GN ATT GGCGG CG	240
GC ACT CAG AG CG CG ACC CGGC CTC GGNGGTG AC GG CGGT GA CG CGG TGAC	290

(2) INFORMATION FOR SEQ ID NO:38:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

GATCCAGTGG CATGGNGGGT GTCAGTGGAA GCAT

34

## (2) INFORMATION FOR SEQ ID NO:39:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 155 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

GATCGCTGCT CGTCCCCCCC TTGCCGCCGA CGCCACCGGT CCCACCGTTA CCGAACAAAGC 60

TGGCGTGGTC GCCAGCACCC CGGGCACCGC CGACGCCGGA GTCGAACAAAT GGCACCGTCG 120

TATCCCCACC ATTGCCGCCG GNCCCACCGG CACCG 155

## (2) INFORMATION FOR SEQ ID NO:40:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 53 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

ATGGCGTTCA CGGGGCGCCG GGGACCGGGC AGCCCGGNNG GGCCGGGGGG TGG 53

## (2) INFORMATION FOR SEQ ID NO:41:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 132 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GATCCACCGC GGGTGCAGAC GGTGCCCGCG GCGCCACCCC GACCAGCGGC GGCAACGGCG 60

GCACCGGCGG CAACGGCGCG AACGCCACCG TCGTCGGNNG GGCCGGCGGG GCCGGCGGCA 120

AGGGCGGCAA CG 132

## (2) INFORMATION FOR SEQ ID NO:42:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 132 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

GATCGGGCGC CGGNACGGNC GGGGACGGCG GCAAGGGCGG NAACGGGGGC GCCGNAGCCA 60

CCNGCCAAGA ATCCTCCGNG TCCNCCAATG GCGCGAATGG CGGACAGGGC GGCAACGGCG 120

GCANCGCGG CA 132

## (2) INFORMATION FOR SEQ ID NO:43:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 702 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

CGGCACGAGG ATCGGTACCC CGCGGCATCG GCAGCTGCCG ATTGCCGGG TTTCCCCACC	60
CGAGGAAAGC CGCTACCAGA TGGCGCTGCC GAAGTAGGGC GATCCGTTCG CGATGCCGGC	120
ATGAACGGGC GGCATCAAAT TAGTGCAGGA ACCTTCAGT TTAGCGACGA TAATGGCTAT	180
AGCACTAAGG AGGATGATCC GATATGACGC AGTCGCAGAC CGTGACGGTG GATCAGCAAG	240
AGATTTGAA CAGGGCCAAC GAGGTGGAGG CCCCAGATGGC GGACCCACCG ACTGATGTCC	300
CCATCACACC GTGCGAACTC ACGGNGGNTA AAAACGCCGC CCAACAGNTG GTNTTGTCCG	360
CCGACAACAT GCGGGAATAC CTGGCGGCCG GTGCCAAAGA GCGGCAGCGT CTGGCGACCT	420
CGCTGCGCAA CGCGGCCAAG GNGTATGGCG AGGTTGATGA GGAGGCTGCG ACCGCGCTGG	480
ACAACGACGG CGAAGGAACT GTGCAGGCAG AATCGGCCGG GGCGTCGGA GGGGACAGTT	540
CGGCCGAACT AACCGATACG CCGAGGGTGG CCACGGCCGG TGAACCCAAC TTCATGGATC	600
TCAAAGAACG GGCAAGGAAG CTCGAAACGG GCGACCAAGG CGCATCGCTC GCGCACTGNG	660
GGGATGGGTG GAACACTTNC ACCCTGACGC TGCAAGGCGA CG	702

## (2) INFORMATION FOR SEQ ID NO:44:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 298 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

GAAGCCGCAG CGCTGTCGGG CGACGTGGCG GTCAAAGCGG CATCGCTCGG TGGCGGTGGA	60
GGCGGCCGGGG TGCCGTCGGC GCCGTTGGGA TCCGCGATCG GGGGCGCCGA ATCGGTGCGG	120
CCCGCTGGCG CTGGTGACAT TGCCGGCTTA GGCCAGGGAA GGGCCGGCGG CGGCGCCGCG	180
CTGGGCGGCG GTGGCATGGG AATGCCGATG GGTGCCGCGC ATCAGGGACA AGGGGGCGCC	240
AAGTCCAAGG GTTCTCAGCA GGAAGACGAG GCGCTCTACA CCGAGGATCC TCGTGCCG	298

## (2) INFORMATION FOR SEQ ID NO:45:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1058 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

CGGCACGAGG ATCGAATCGC GTGCCGGGA GCACAGCGTC GCACTGCACC AGTGGAGGAG	60
CCATGACCTA CTCGCCGGGT AACCCCGGAT ACCCGCAAGC GCAGCCCGCA GGCTCCTACG	120
GAGGCACGTAC ACCCTCGTTC GCCCACGCGC ATGAGGGTGC GAGCAAGCTA CCGATGTACC	180
TGAACATCGC GGTGGCAGTG CTCGGTCTGG CTGCGTACTT CGCCAGCTTC GGCCCAATGT	240
TCACCCCTCAG TACCGAACTC GGGGGGGGTG ATGGCGCAGT GTCCGGTGAC ACTGGGCTGC	300
CGGTGCGGGGT GGCTCTGCTG GCTGCGCTGC TTGCCGGGGT GGTTCTGGTG CCTAAGGCCA	360
AGAGCCATGT GACGGTAGTT GCGGTGCTCG GGGTACTCGG CGTATTCCTG ATGGTCTCGG	420

CGACGTTAA CAAGCCCAGC GCCTATTGA CCGGTTGGC ATTGTGGTT GTGTTGGCTT	480
TCATCGTGTT CCAGGCGGTT GCGGCAGTCC TGGCGCTCTT GGTGGAGACC GGCGCTATCA	540
CCGCGCCGGC GCCGCGGCC AAGTCGACC CGTATGGACA GTACGGCGG TACGGGCAGT	600
ACGGGCAGTA CGGGGTGCAG CCGGGTGGGT ACTACGGTCA GCAGGGTGCT CAGCAGGCCG	660
CGGGACTGCA GTCGCCCCGC CCGCAGCAGT CTCCGCAGCC TCCCGGATAT GGGTCGCAGT	720
ACGGCGGCTA TTCTGTCAGT CCGAGCCAAT CGGGCAGTGG ATACACTGCT CAGCCCCCGG	780
CCCAGCCGCC GGCGCAGTCC GGGTCGCAAC AATCGCACCA GGGCCCATCC ACGCCACCTA	840
CCGGCTTCC GAGCTTCAGC CCACCAACAC CGGTCACTGC CGGGACGGGG TCGCAGGCTG	900
GTTCGGCTCC AGTCAACTAT TCAAACCCA GCGGGGGCGA GCAGTCGTG TCCCCCGGGG	960
GGGCGCCGGT CTAACCGGGC GTTCCCGCGT CCGGTGCGC GTGTGCGCGA AGAGTGAACA	1020
GGGTGTCAGC AAGCGCGGAC GATCCTCGTG CCGAATT	1058

## (2) INFORMATION FOR SEQ ID NO:46:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 327 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

CGGCACGAGA GACCGATGCC GCTACCCCTCG CGCAGGAGGC AGGTAATTTC GAGCGGATCT	60
CCGGCGACCT GAAAACCCAG ATCGACCAGG TGGAGTCGAC GGCAGGTTCG TTGCAGGGCC	120
AGTGGCGCGG CGCGGCGGGG ACGGCCGCC AGGCCGCGGT GGTGCGCTTC CAAGAACGAG	180
CCAATAAGCA GAAGCAGGAA CTCGACGAGA TCTCGACGAA TATTGTCAG GCCGGCGTCC	240
AATACTCGAG GGCCGACGAG GAGCAGCAGC AGGCCTGTC CTCGCAAATG GGCTTCTGAC	300

CCGCTAATAC GAAAAGAAC GGAGCAA 327

(2) INFORMATION FOR SEQ ID NO:47:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 170 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

CGGTCGCGAT GATGGCGTTG TCGAACGTGA CCGATTCTGT ACCGCCGTG TTGAGATCAA 60  
CCAACAAACGT GTTGGCGTCG GCAAATGTGC CGNACCCGTG GATCTCGGTG ATCTTGTCT 120  
TCTTCATCAG GAAGTGCACA CGGGCCACCC TGCCCTCGGN TACCTTTCGG 170

(2) INFORMATION FOR SEQ ID NO:48:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 127 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

GATCCGGCGG CACGGGGGGT GCCGGCGGCA GCACCGCTGG CGCTGGCGGC AACGGCGGGG 60  
CCGGGGGGTGG CGGCAGAACCC GGTGGGTTGC TCTTCGGCAA CGGCAGGTGCC GGCGGGCACG 120  
GGGCCGT 127

(2) INFORMATION FOR SEQ ID NO:49:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 81 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

CGGCGGGCAAG GGCGGCACCG CGGGCACCGG GAGCGGCGCG GCCGGCGGCA ACGGCGGCAA	60
CGGCGGCTCC GGCCTCAACG G	81

## (2) INFORMATION FOR SEQ ID NO:50:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 149 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

GATCAGGGCT GGCCGGCTCC GGCCAGAAGG GCGGTAACGG AGGAGCTGCC GGATTGTTTG	60
GCAACGGCGG GGCCGGNGGT GCCGGCGCGT CCAACCAAGC CGGTAACGGC GGNGCCGGCG	120
GAAACGGTGG TGCCGGTGGG CTGATCTGG	149

## (2) INFORMATION FOR SEQ ID NO:51:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 355 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

CGGCACGAGA TCACACCTAC CGAGTGATCG AGATCGTCGG GACCTCGCCC GACGGTGTG	60
ACGCCGNAAT CCAGGGCGGT CTGGCCCGAG CTGCGCAGAC CATGCGCGCG CTGGACTGGT	120
TCGAAGTACA GTCAATTGCA GGCCACCTGG TCGACGGAGC GGTCGCGCAC TTCCAGGTGA	180
CTATGAAAGT CGGCTTCCGC CTGGAGGATT CCTGAACCTT CAAGCGCGC CGATAACTGA	240
GGTGCATCAT TAAGCGACTT TTCCAGAACCA TCCTGACGCG CTCGAAACGC GGTTCAGCCG	300
ACGGTGGCTC CGCCGAGGCG CTGCCTCCAA AATCCCTGCG ACAATTGTC GGC GG	355

## (2) INFORMATION FOR SEQ ID NO:52:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 999 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

ATGCATCACC ATCACCATCA CATGCATCAG GTGGACCCCCA ACTTGACACG TCGCAAGGGA	60
CGATTGGCGG CACTGGCTAT CGCGGCGATG GCCAGCGCCA GCCTGGTGAC CGTTGCGGTG	120
CCCGCGACCG CCAACGCCGA TCCGGAGCCA GCGCCCCCGG TACCCACAAC GGCCGCCTCG	180
CCGCCGTCGA CCCGCTGCCAGC GCCACCCGCA CGGGCGACAC CTGTTGCCCC CCCACCAACG	240
GCCGCCGCCA ACACGCCGAA TGCCCAGCCG GGCGATCCCA ACGCAGCACC TCCGCCGGCC	300
GACCCGAACG CACCGCCGCC ACCTGTCATT GCCCCAAACG CACCCCAACC TGTCCGGATC	360
GACAACCCGG TTGGAGGATT CAGCTTCGCG CTGCCTGCTG GCTGGGTGGA GTCTGACGCC	420

GCCCACCTCG ACTACGGTTC AGCACTCCTC AGCAAAACCA CCGGGGACCC GCCATTTCCC	480
GGACAGCCGC CGCCGGTGGC CAATGACACC CGTATCGTGC TCGGCCGGCT AGACCAAAAG	540
CTTACGCCA GCGCCGAAGC CACCGACTCC AAGGCCGC GG CCCGGTTGGG CTCGGACATG	600
GGTGAGTTCT ATATGCCCTA CCCGGGCACC CGGATCAACC AGGAAACCGT CTCGCTCGAC	660
GCCAACGGGG TGTCTGGAAG CGCGTCGTAT TACGAAGTCA AGTTCAGCGA TCCGAGTAAG	720
CCGAACGGCC AGATCTGGAC GGGCGTAATC GGCTCGCCCG CGGCGAACGC ACCGGACGCC	780
GGGGCCCCCTC AGCGCTGGTT TGTGGTATGG CTCGGGACCG CCAACAACCC GGTGGACAAG	840
GGCGCGGCCA AGGCGCTGGC CGAATCGATC CGGCCTTGG TCGCCCCGCC GCCGGCGCCG	900
GCACCGGCTC CTGCAGAGCC CGCTCCGGCG CGGGCGCCGG CCGGGGAAAGT CGCTCCTACC	960
CCGACGACAC CGACACCGCA GCGGACCTTA CGGGCTGA	999

## (2) INFORMATION FOR SEQ ID NO:53:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 332 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

Met His His His His His Met His Glu Val Asp Pro Asn Leu Thr	
1 5 10 15	
Arg Arg Lys Gly Arg Leu Ala Ala Leu Ala Ile Ala Ala Met Ala Ser	
20 25 30	
Ala Ser Leu Val Thr Val Ala Val Pro Ala Thr Ala Asn Ala Asp Pro	
35 40 45	
Glu Pro Ala Pro Pro Val Pro Thr Thr Ala Ala Ser Pro Pro Ser Thr	
50 55 60	

Ala Ala Ala Pro Pro Ala Pro Ala Thr Pro Val Ala Pro Pro Pro Pro  
65 70 75 80

Ala Ala Ala Asn Thr Pro Asn Ala Gln Pro Gly Asp Pro Asn Ala Ala  
85 90 95

Pro Pro Pro Ala Asp Pro Asn Ala Pro Pro Pro Pro Val Ile Ala Pro  
100 105 110

Asn Ala Pro Gln Pro Val Arg Ile Asp Asn Pro Val Gly Gly Phe Ser  
115 120 125

Phe Ala Leu Pro Ala Gly Trp Val Glu Ser Asp Ala Ala His Phe Asp  
130 135 140

Tyr Gly Ser Ala Leu Leu Ser Lys Thr Thr Gly Asp Pro Pro Phe Pro  
145 150 155 160

Gly Gln Pro Pro Pro Val Ala Asn Asp Thr Arg Ile Val Leu Gly Arg  
165 170 175

Leu Asp Gln Lys Leu Tyr Ala Ser Ala Glu Ala Thr Asp Ser Lys Ala  
180 185 190

Ala Ala Arg Leu Gly Ser Asp Met Gly Glu Phe Tyr Met Pro Tyr Pro  
195 200 205

Gly Thr Arg Ile Asn Gln Glu Thr Val Ser Leu Asp Ala Asn Gly Val  
210 215 220

Ser Gly Ser Ala Ser Tyr Tyr Gln Val Lys Phe Ser Asp Pro Ser Lys  
225 230 235 240

Pro Asn Gly Gln Ile Trp Thr Gly Val Ile Gly Ser Pro Ala Ala Asn  
245 250 255

Ala Pro Asp Ala Gly Pro Pro Gln Arg Trp Phe Val Val Trp Leu Gly  
260 265 270

Thr Ala Asn Asn Pro Val Asp Lys Gly Ala Ala Lys Ala Leu Ala Glu  
275 280 285

Ser Ile Arg Pro Leu Val Ala Pro Pro Pro Ala Pro Ala Pro Ala Pro  
290 295 300

Ala Glu Pro Ala Pro Ala Pro Ala Pro Ala Gly Glu Val Ala Pro Thr  
305                   310                   315                   320  
Pro Thr Thr Pro Thr Pro Gln Arg Thr Leu Pro Ala  
325                   330

## (2) INFORMATION FOR SEQ ID NO:54:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

Asp Pro Val Asp Ala Val Ile Asn Thr Thr Xaa Asn Tyr Gly Gln Val  
1                   5                   10                   15  
Val Ala Ala Leu  
20

## (2) INFORMATION FOR SEQ ID NO:55:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 15 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

Ala Val Glu Ser Gly Met Leu Ala Leu Gly Thr Pro Ala Pro Ser  
1                   5                   10                   15

## (2) INFORMATION FOR SEQ ID NO:56:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

Ala Ala Met Lys Pro Arg Thr Gly Asp Gly Pro Leu Glu Ala Ala Lys  
1               5                           10                           15  
Glu Gly Arg

## (2) INFORMATION FOR SEQ ID NO:57:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

Tyr Tyr Trp Cys Pro Gly Gin Pro Phe Asp Pro Ala Trp Gly Pro  
1               5                           10                           15

## (2) INFORMATION FOR SEQ ID NO:58:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

Asp Ile Gly Ser Glu Ser Thr Glu Asp Gln Gln Xaa Ala Val  
1 5 10

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

Ala Glu Glu Ser Ile Ser Thr Xaa Glu Xaa Ile Val Pro  
1 5 10

(2) INFORMATION FOR SEO ID NO: 60:

(i) SEQUENCE CHARACTERISTICS:

- SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 17 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

Asp Pro Glu Pro Ala Pro Pro Val Pro Thr Ala Ala Ala Ala Pro Pro  
1 5 10 15

A1a

## (2) INFORMATION FOR SEQ ID NO:61:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 15 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

Ala Pro Lys Thr Tyr Xaa Glu Glu Leu Lys Gly Thr Asp Thr Gly  
1               5                   10                   15

## (2) INFORMATION FOR SEQ ID NO:62:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 30 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

Asp Pro Ala Ser Ala Pro Asp Val Pro Thr Ala Ala Gln Gln Thr Ser  
1               5                   10                   15

Leu Leu Asn Asn Leu Ala Asp Pro Asp Val Ser Phe Ala Asp  
20               25                   30

## (2) INFORMATION FOR SEQ ID NO:63:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 187 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

Thr Gly Ser Leu Asn Gln Thr His Asn Arg Arg Ala Asn Glu Arg Lys  
1 5 10 15

Asn Thr Thr Met Lys Met Val Lys Ser Ile Ala Ala Gly Leu Thr Ala  
20 25 30

Ala Ala Ala Ile Gly Ala Ala Ala Gly Val Thr Ser Ile Met Ala  
35 40 45

Gly Gly Pro Val Val Tyr Gln Met Gln Pro Val Val Phe Gly Ala Pro  
50 55 60

Leu Pro Leu Asp Pro Ala Ser Ala Pro Asp Val Pro Thr Ala Ala Gln  
65 70 75 80

Leu Thr Ser Leu Leu Asn Ser Leu Ala Asp Pro Asn Val Ser Phe Ala  
85 90 95

Asn Lys Gly Ser Leu Val Glu Gly Ile Gly Gly Thr Glu Ala Arg  
100 105 110

Ile Ala Asp His Lys Leu Lys Lys Ala Ala Glu His Gly Asp Leu Pro  
115 120 125

Leu Ser Phe Ser Val Thr Asn Ile Gln Pro Ala Ala Gly Ser Ala  
130 135 140

Thr Ala Asp Val Ser Val Ser Gly Pro Lys Leu Ser Ser Pro Val Thr  
145 150 155 160

Gln Asn Val Thr Phe Val Asn Gln Gly Gly Trp Met Leu Ser Arg Ala  
165 170 175

Ser Ala Met Glu Leu Leu Gln Ala Ala Gly Xaa  
180 185

## (2) INFORMATION FOR SEQ ID NO:64:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 148 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

Asp	Glu	Val	Thr	Val	Glu	Thr	Thr	Ser	Val	Phe	Arg	Ala	Asp	Phe	Leu
1				5				10						15	
Ser	Glu	Leu	Asp	Ala	Pro	Ala	Gln	Ala	Gly	Thr	Glu	Ser	Ala	Val	Ser
					20			25						30	
Gly	Val	Glu	Gly	Leu	Pro	Pro	Gly	Ser	Ala	Leu	Leu	Val	Val	Lys	Arg
				35			40						45		
Gly	Pro	Asn	Ala	Gly	Ser	Arg	Phe	Leu	Leu	Asp	Gln	Ala	Ile	Thr	Ser
				50			55						60		
Ala	Gly	Arg	His	Pro	Asp	Ser	Asp	Ile	Phe	Leu	Asp	Asp	Val	Thr	Val
				65			70						75		80
Ser	Arg	Arg	His	Ala	Glu	Phe	Arg	Leu	Glu	Asn	Asn	Glu	Phe	Asn	Val
					85			90					95		
Val	Asp	Val	Gly	Ser	Leu	Asn	Gly	Thr	Tyr	Val	Asn	Arg	Glu	Pro	Val
					100			105					110		
Asp	Ser	Ala	Val	Leu	Ala	Asn	Gly	Asp	Glu	Val	Gln	Ile	Gly	Lys	Leu
					115			120					125		
Arg	Leu	Val	Phe	Leu	Thr	Gly	Pro	Lys	Gln	Gly	Glu	Asp	Asp	Gly	Ser
					130			135					140		
Thr	Gly	Gly	Pro												
				145											

(2) INFORMATION FOR SEQ ID NO:65:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 230 amino acids

- (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

Thr Ser Asn Arg Pro Ala Arg Arg Gly Arg Arg Ala Pro Arg Asp Thr  
1 5 10 15

Gly Pro Asp Arg Ser Ala Ser Leu Ser Leu Val Arg His Arg Arg Gln  
20 25 30

Gln Arg Asp Ala Leu Cys Leu Ser Ser Thr Gln Ile Ser Arg Gln Ser  
35 40 45

Asn Leu Pro Pro Ala Ala Gly Gly Ala Ala Asn Tyr Ser Arg Arg Asn  
50 55 60

Phe Asp Val Arg Ile Lys Ile Phe Met Leu Val Thr Ala Val Val Leu  
65                   70                   75                   80

Leu Cys Cys Ser Gly Val Ala Thr Ala Ala Pro Lys Thr Tyr Cys Glu  
85                    90                    95

Glu Leu Lys Gly Thr Asp Thr Gly Gln Ala Cys Gln Ile Gln Met Ser  
100 105 110

Asp Pro Ala Tyr Asn Ile Asn Ile Ser Leu Pro Ser Tyr Tyr Pro Asp  
115 120 125

Gln Lys Ser Leu Glu Asn Tyr Ile Ala Gln Thr Arg Asp Lys Phe Leu  
130 135 140

Ser Ala Ala Thr Ser Ser Thr Pro Arg Glu Ala Pro Tyr Glu Leu Asn  
145 150 155 160

Ile Thr Ser Ala Thr Tyr Gln Ser Ala Ile Pro Pro Arg Gly Thr Gln  
165 170 175

Ala Val Val Leu Xaa Val Tyr His Asn Ala Gly Gly Thr His Pro Thr  
180 185 190

Thr Thr Tyr Lys Ala Phe Asp Trp Asp Gln Ala Tyr Arg Lys Pro Ile  
195 200 205

Thr Tyr Asp Thr Leu Trp Gln Ala Asp Thr Asp Pro Leu Pro Val Val  
210 215 220

Phe Pro Ile Val Ala Arg  
225 230

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 132 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

Thr Ala Ala Ser Asp Asn Phe Gln Leu Ser Gln Gly Gly Gln Gly Phe  
1 5 10 15

Ala Ile Pro Ile Gly Gln Ala Met Ala Ile Ala Gly Gln Ile Arg Ser  
20 25 30

Gly Gly Gly Ser Pro Thr Val His Ile Gly Pro Thr Ala Phe Leu Gly  
35 40 45

Leu Gly Val Val Asp Asn Asn Gly Asn Gly Ala Arg Val Gln Arg Val  
50 55 60

Val Gly Ser Ala Pro Ala Ala Ser Leu Gly Ile Ser Thr Gly Asp Val  
65 70 75 80

Ile Thr Ala Val Asp Gly Ala Pro Ile Asn Ser Ala Thr Ala Met Ala  
85 90 95

Asp Ala Leu Asn Gly His His Pro Gly Asp Val Ile Ser Val Asn Trp  
100 105 110

Gln Thr Lys Ser Gly Gly Thr Arg Thr Gly Asn Val Thr Leu Ala Glu  
115 120 125  
Gly Pro Pro Ala  
130

## (2) INFORMATION FOR SEQ ID NO:67:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 100 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

Val Pro Leu Arg Ser Pro Ser Met Ser Pro Ser Lys Cys Leu Ala Ala  
1 5 10 15

Ala Gln Arg Asn Pro Val Ile Arg Arg Arg Arg Leu Ser Asn Pro Pro  
20 25 30

Pro Arg Lys Tyr Arg Ser Met Pro Ser Pro Ala Thr Ala Ser Ala Gly  
35 40 45

Met Ala Arg Val Arg Arg Ala Ile Trp Arg Gly Pro Ala Thr Xaa  
50 55 60

Ser Ala Gly Met Ala Arg Val Arg Arg Trp Xaa Val Met Pro Xaa Val  
65 70 75 80

Ile Gln Ser Thr Xaa Ile Arg Xaa Xaa Gly Pro Phe Asp Asn Arg Gly  
85 90 95

Ser Glu Arg Lys  
100

## (2) INFORMATION FOR SEQ ID NO:68:

- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 163 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

Met Thr Asp Asp Ile Leu Leu Ile Asp Thr Asp Glu Arg Val Arg Thr  
1 5 10 15

Leu Thr Leu Asn Arg Pro Gln Ser Arg Asn Ala Leu Ser Ala Ala Leu  
20 25 30

Arg Asp Arg Phe Phe Ala Xaa Leu Xaa Asp Ala Glu Xaa Asp Asp Asp  
35 40 45

Ile Asp Val Val Ile Leu Thr Gly Ala Asp Pro Val Phe Cys Ala Gly  
50 55 60

Leu Asp Leu Lys Val Ala Gly Arg Ala Asp Arg Ala Ala Gly His Leu  
65 70 75 80

Thr Ala Val Gly Gly His Asp Gln Ala Gly Asp Arg Arg Asp Gln Arg  
85 90 95

Arg Arg Gly His Arg Arg Ala Arg Thr Gly Ala Val Leu Arg His Pro  
100 105 110

Asp Arg Leu Arg Ala Arg Pro Leu Arg Arg His Pro Arg Pro Gly Gly  
115 120 125

Ala Ala Ala His Leu Gly Thr Gln Cys Val Leu Ala Ala Lys Gly Arg  
130 135 140

His Arg Xaa Gly Pro Val Asp Glu Pro Asp Arg Arg Leu Pro Val Arg  
145 150 155 160

Asp Arg Arg

(2) INFORMATION FOR SEQ ID NO:69:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 344 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

Met Lys Phe Val Asn His Ile Glu Pro Val Ala Pro Arg Arg Ala Gly  
1 5 10 15

Gly Ala Val Ala Glu Val Tyr Ala Glu Ala Arg Arg Glu Phe Gly Arg  
20 25 30

Leu Pro Glu Pro Leu Ala Met Leu Ser Pro Asp Glu Gly Leu Leu Thr  
35 40 45

Ala Gly Trp Ala Thr Leu Arg Glu Thr Leu Leu Val Gly Gln Val Pro  
50 55 60

Arg Gly Arg Lys Glu Ala Val Ala Ala Val Ala Ala Ser Leu Arg  
65 70 75 80

Cys Pro Trp Cys Val Asp Ala His Thr Thr Met Leu Tyr Ala Ala Gly  
85 90 95

Gln Thr Asp Thr Ala Ala Ala Ile Leu Ala Gly Thr Ala Pro Ala Ala  
100 105 110

Gly Asp Pro Asn Ala Pro Tyr Val Ala Trp Ala Ala Gly Thr Gly Thr  
115 120 125

Pro Ala Gly Pro Pro Ala Pro Phe Gly Pro Asp Val Ala Ala Glu Tyr  
130 135 140

Leu Gly Thr Ala Val Gln Phe His Phe Ile Ala Arg Leu Val Leu Val  
145 150 155 160

Leu Leu Asp Glu Thr Phe Leu Pro Gly Gly Pro Arg Ala Gln Gln Leu  
165 170 175

100

Met Arg Arg Ala Gly Gly Leu Val Phe Ala Arg Lys Val Arg Ala Glu  
180 185 190

His Arg Pro Gly Arg Ser Thr Arg Arg Leu Glu Pro Arg Thr Leu Pro  
195 200 205

Asp Asp Leu Ala Trp Ala Thr Pro Ser Glu Pro Ile Ala Thr Ala Phe  
210 215 220

Ala Ala Leu Ser His His Leu Asp Thr Ala Pro His Leu Pro Pro Pro  
225 230 235 240

Thr Arg Gln Val Val Arg Arg Val Val Gly Ser Trp His Gly Glu Pro  
245 250 255

Met Pro Met Ser Ser Arg Trp Thr Asn Glu His Thr Ala Glu Leu Pro  
260 265 270

Ala Asp Leu His Ala Pro Thr Arg Leu Ala Leu Leu Thr Gly Leu Ala  
275 280 285

Pro His Gln Val Thr Asp Asp Asp Val Ala Ala Ala Arg Ser Leu Leu  
290 295 300

Asp Thr Asp Ala Ala Leu Val Gly Ala Leu Ala Trp Ala Ala Phe Thr  
305 310 315 320

Ala Ala Arg Arg Ile Gly Thr Trp Ile Gly Ala Ala Ala Glu Gly Gln  
325 330 335

Val Ser Arg Gln Asn Pro Thr Gly  
340

## (2) INFORMATION FOR SEQ ID NO:70:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 485 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

Asp Asp Pro Asp Met Pro Gly Thr Val Ala Lys Ala Val Ala Asp Ala  
1 5 10 15

Leu Gly Arg Gly Ile Ala Pro Val Glu Asp Ile Gln Asp Cys Val Glu  
20 25 30

Ala Arg Leu Gly Glu Ala Gly Leu Asp Asp Val Ala Arg Val Tyr Ile  
35 40 45

Ile Tyr Arg Gln Arg Arg Ala Glu Leu Arg Thr Ala Lys Ala Leu Leu  
50 55 60

Gly Val Arg Asp Glu Leu Lys Leu Ser Leu Ala Ala Val Thr Val Leu  
65 70 75 80

Arg Glu Arg Tyr Leu Leu His Asp Glu Gln Gly Arg Pro Ala Glu Ser  
85 90 95

Thr Gly Glu Leu Met Asp Arg Ser Ala Arg Cys Val Ala Ala Ala Glu  
100 105 110

Asp Gln Tyr Glu Pro Gly Ser Ser Arg Arg Trp Ala Glu Arg Phe Ala  
115 120 125

Thr Leu Leu Arg Asn Leu Glu Phe Leu Pro Asn Ser Pro Thr Leu Met  
130 135 140

Asn Ser Gly Thr Asp Leu Gly Leu Leu Ala Gly Cys Phe Val Leu Pro  
145 150 155 160

Ile Glu Asp Ser Leu Gln Ser Ile Phe Ala Thr Leu Gly Gln Ala Ala  
165 170 175

Glu Leu Gln Arg Ala Gly Gly Thr Gly Tyr Ala Phe Ser His Leu  
180 185 190

Arg Pro Ala Gly Asp Arg Val Ala Ser Thr Gly Gly Thr Ala Ser Gly  
195 200 205

Pro Val Ser Phe Leu Arg Leu Tyr Asp Ser Ala Ala Gly Val Val Ser  
210 215 220

102

Met Gly Gly Arg Arg Arg Gly Ala Cys Met Ala Val Leu Asp Val Ser  
225 230 235 240

His Pro Asp Ile Cys Asp Phe Val Thr Ala Lys Ala Glu Ser Pro Ser  
245 250 255

Glu Leu Pro His Phe Asn Leu Ser Val Gly Val Thr Asp Ala Phe Leu  
260 265 270

Arg Ala Val Glu Arg Asn Gly Leu His Arg Leu Val Asn Pro Arg Thr  
275 280 285

Gly Lys Ile Val Ala Arg Met Pro Ala Ala Glu Leu Phe Asp Ala Ile  
290 295 300

Cys Lys Ala Ala His Ala Gly Gly Asp Pro Gly Leu Val Phe Leu Asp  
305 310 315 320

Thr Ile Asn Arg Ala Asn Pro Val Pro Gly Arg Gly Arg Ile Glu Ala  
325 330 335

Thr Asn Pro Cys Gly Glu Val Pro Leu Leu Pro Tyr Glu Ser Cys Asn  
340 345 350

Leu Gly Ser Ile Asn Leu Ala Arg Met Leu Ala Asp Gly Arg Val Asp  
355 360 365

Trp Asp Arg Leu Glu Glu Val Ala Gly Val Ala Val Arg Phe Leu Asp  
370 375 380

Asp Val Ile Asp Val Ser Arg Tyr Pro Phe Pro Glu Leu Gly Glu Ala  
385 390 395 400

Ala Arg Ala Thr Arg Lys Ile Gly Leu Gly Val Met Gly Leu Ala Glu  
405 410 415

Leu Leu Ala Ala Leu Gly Ile Pro Tyr Asp Ser Glu Glu Ala Val Arg  
420 425 430

Leu Ala Thr Arg Leu Met Arg Arg Ile Gln Gln Ala Ala His Thr Ala  
435 440 445

Ser Arg Arg Leu Ala Glu Glu Arg Gly Ala Phe Pro Ala Phe Thr Asp  
450 455 460

103

Ser Arg Phe Ala Arg Ser Gly Pro Arg Arg Asn Ala Gln Val Thr Ser  
465                    470                    475                    480  
Val Ala Pro Thr Gly  
485

## (2) INFORMATION FOR SEQ ID NO:71:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 267 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

Gly Val Ile Val Leu Asp Leu Glu Pro Arg Gly Pro Leu Pro Thr Glu  
1                5                10                15  
Ile Tyr Trp Arg Arg Arg Gly Leu Ala Leu Gly Ile Ala Val Val Val  
20                25                30  
Val Gly Ile Ala Val Ala Ile Val Ile Ala Phe Val Asp Ser Ser Ala  
35                40                45  
Gly Ala Lys Pro Val Ser Ala Asp Lys Pro Ala Ser Ala Gln Ser His  
50                55                60  
Pro Gly Ser Pro Ala Pro Gln Ala Pro Gln Pro Ala Gly Gln Thr Glu  
65                70                75                80  
Gly Asn Ala Ala Ala Ala Pro Pro Gln Gly Gln Asn Pro Glu Thr Pro  
85                90                95  
Thr Pro Thr Ala Ala Val Gln Pro Pro Pro Val Leu Lys Glu Gly Asp  
100                105                110  
Asp Cys Pro Asp Ser Thr Leu Ala Val Lys Gly Leu Thr Asn Ala Pro  
115                120                125  
Gln Tyr Tyr Val Gly Asp Gln Pro Lys Phe Thr Met Val Val Thr Asn

104

130

135

140

Ile Gly Leu Val Ser Cys Lys Arg Asp Val Gly Ala Ala Val Leu Ala  
145                    150                    155                    160

Ala Tyr Val Tyr Ser Leu Asp Asn Lys Arg Leu Trp Ser Asn Leu Asp  
165                    170                    175

Cys Ala Pro Ser Asn Glu Thr Leu Val Lys Thr Phe Ser Pro Gly Glu  
180                    185                    190

Gln Val Thr Thr Ala Val Thr Trp Thr Gly Met Gly Ser Ala Pro Arg  
195                    200                    205

Cys Pro Leu Pro Arg Pro Ala Ile Gly Pro Gly Thr Tyr Asn Leu Val  
210                    215                    220

Val Gln Leu Gly Asn Leu Arg Ser Leu Pro Val Pro Phe Ile Leu Asn  
225                    230                    235                    240

Gln Pro Pro Pro Pro Gly Pro Val Pro Ala Pro Gly Pro Ala Gln  
245                    250                    255

Ala Pro Pro Pro Glu Ser Pro Ala Gln Gly Gly  
260                    265

(2) INFORMATION FOR SEQ ID NO:72:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 97 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

Leu Ile Ser Thr Gly Lys Ala Ser His Ala Ser Leu Gly Val Gln Val  
1                    5                    10                    15

Thr Asn Asp Lys Asp Thr Pro Gly Ala Lys Ile Val Glu Val Val Ala  
20                    25                    30

105

Gly Gly Ala Ala Ala Asn Ala Gly Val Pro Lys Gly Val Val Val Thr  
35 40 45

Lys Val Asp Asp Arg Pro Ile Asn Ser Ala Asp Ala Leu Val Ala Ala  
50 55 60

Val Arg Ser Lys Ala Pro Gly Ala Thr Val Ala Leu Thr Phe Gln Asp  
65 70 75 80

Pro Ser Gly Gly Ser Arg Thr Val Gln Val Thr Leu Gly Lys Ala Glu  
85 90 95

Gln

## (2) INFORMATION FOR SEQ ID NO:73:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 364 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

Gly Ala Ala Val Ser Leu Leu Ala Ala Gly Thr Leu Val Leu Thr Ala  
1 5 10 15

Cys Gly Gly Gly Thr Asn Ser Ser Ser Ser Gly Ala Gly Gly Thr Ser  
20 25 30

Gly Ser Val His Cys Gly Gly Lys Lys Glu Leu His Ser Ser Gly Ser  
35 40 45

Thr Ala Gln Glu Asn Ala Met Glu Gln Phe Val Tyr Ala Tyr Val Arg  
50 55 60

Ser Cys Pro Gly Tyr Thr Leu Asp Tyr Asn Ala Asn Gly Ser Gly Ala  
65 70 75 80

106

Gly Val Thr Gln Phe Leu Asn Asn Glu Thr Asp Phe Ala Gly Ser Asp  
85 90 95

Val Pro Leu Asn Pro Ser Thr Gly Gln Pro Asp Arg Ser Ala Glu Arg  
100 105 110

Cys Gly Ser Pro Ala Trp Asp Leu Pro Thr Val Phe Gly Pro Ile Ala  
115 120 125

Ile Thr Tyr Asn Ile Lys Gly Val Ser Thr Leu Asn Leu Asp Gly Pro  
130 135 140

Thr Thr Ala Lys Ile Phe Asn Gly Thr Ile Thr Val Trp Asn Asp Pro  
145 150 155 160

Gln Ile Gln Ala Leu Asn Ser Gly Thr Asp Leu Pro Pro Thr Pro Ile  
165 170 175

Ser Val Ile Phe Arg Ser Asp Lys Ser Gly Thr Ser Asp Asn Phe Gln  
180 185 190

Lys Tyr Leu Asp Gly Val Ser Asn Gly Ala Trp Gly Lys Gly Ala Ser  
195 200 205

Glu Thr Phe Ser Gly Gly Val Gln Val Gln Ala Ser Gly Asn Asn Gly  
210 215 220

Thr Ser Ala Leu Leu Gln Thr Thr Asp Gly Ser Ile Thr Tyr Asn Glu  
225 230 235 240

Trp Ser Phe Ala Val Gln Lys Gln Leu Asn Met Ala Gln Ile Ile Thr  
245 250 255

Ser Ala Gly Pro Asp Pro Val Ala Ile Thr Thr Glu Ser Val Gly Lys  
260 265 270

Thr Ile Ala Gly Ala Lys Ile Met Gly Gln Gly Asn Asp Leu Val Leu  
275 280 285

Asp Thr Ser Ser Phe Tyr Arg Pro Thr Gln Pro Gly Ser Tyr Pro Ile  
290 295 300

Val Leu Ala Thr Tyr Glu Ile Val Cys Ser Lys Tyr Pro Asp Ala Thr  
305 310 315 320

Thr Gly Thr Ala Val Arg Ala Phe Met Gln Ala Ala Ile Gly Pro Gly  
 325 330 335

Gln Glu Gly Leu Asp Gln Tyr Gly Ser Ile Pro Leu Pro Lys Ser Phe  
 340 345 350

Gln Ala Lys Leu Ala Ala Ala Val Asn Ala Ile Ser  
 355 360

(2) INFORMATION FOR SEQ ID NO:74:

### (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 309 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

Gln	Ala	Ala	Ala	Gly	Arg	Ala	Val	Arg	Arg	Thr	Gly	His	Ala	Glu	Asp
1				5						10				15	
Gln	Thr	His	Gln	Asp	Arg	Leu	His	His	Gly	Cys	Arg	Arg	Ala	Ala	Val
						20				25				30	
Val	Val	Arg	Gln	Asp	Arg	Ala	Ser	Val	Ser	Ala	Thr	Ser	Ala	Arg	Pro
						35			40				45		
Pro	Arg	Arg	His	Pro	Ala	Gln	Gly	His	Arg	Arg	Arg	Val	Ala	Pro	Ser
						50			55				60		
Gly	Gly	Arg	Arg	Arg	Pro	His	Pro	His	His	Val	Gln	Pro	Asp	Asp	Arg
						65			70			75			80
Arg	Asp	Arg	Pro	Ala	Leu	Leu	Asp	Arg	Thr	Gln	Pro	Ala	Glu	His	Pro
									85				90		95
Asp	Pro	His	Arg	Arg	Gly	Pro	Ala	Asp	Pro	Gly	Arg	Val	Arg	Gly	Arg
						100				105				110	
Gly	Arg	Leu	Arg	Arg	Val	Asp	Asp	Gly	Arg	Leu	Gln	Pro	Asp	Arg	Asp

108

115	120	125
Ala Asp His Gly Ala Pro Val Arg Gly Arg Gly Pro His Arg Gly Val		
130	135	140
Gln His Arg Gly Gly Pro Val Phe Val Arg Arg Val Pro Gly Val Arg		
145	150	155
160		
Cys Ala His Arg Arg Gly His Arg Arg Val Ala Ala Pro Gly Gln Gly		
165	170	175
Asp Val Leu Arg Ala Gly Leu Arg Val Glu Arg Leu Arg Pro Val Ala		
180	185	190
Ala Val Glu Asn Leu His Arg Gly Ser Gln Arg Ala Asp Gly Arg Val		
195	200	205
Phe Arg Pro Ile Arg Arg Gly Ala Arg Leu Pro Ala Arg Arg Ser Arg		
210	215	220
Ala Gly Pro Gln Gly Arg Leu His Leu Asp Gly Ala Gly Pro Ser Pro		
225	230	235
240		
Leu Pro Ala Arg Ala Gly Gln Gln Pro Ser Ser Ala Gly Gly Arg		
245	250	255
Arg Ala Gly Gly Ala Glu Arg Ala Asp Pro Gly Gln Arg Gly Arg His		
260	265	270
His Gln Gly Gly His Asp Pro Gly Arg Gln Gly Ala Gln Arg Gly Thr		
275	280	285
Ala Gly Val Ala His Ala Ala Gly Pro Arg Arg Ala Ala Val Arg		
290	295	300
Asn Arg Pro Arg Arg		
305		

## (2) INFORMATION FOR SEQ ID NO:75:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 580 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

Ser Ala Val Trp Cys Leu Asn Gly Phe Thr Gly Arg His Arg His Gly  
1 5 10 15

Arg Cys Arg Val Arg Ala Ser Gly Trp Arg Ser Ser Asn Arg Trp Cys  
20 25 30

Ser Thr Thr Ala Asp Cys Cys Ala Ser Lys Thr Pro Thr Gln Ala Ala  
35 40 45

Ser Pro Leu Glu Arg Arg Phe Thr Cys Cys Ser Pro Ala Val Gly Cys  
50 55 60

Arg Phe Arg Ser Phe Pro Val Arg Arg Leu Ala Leu Gly Ala Arg Thr  
65 70 75 80

Ser Arg Thr Leu Gly Val Arg Arg Thr Leu Ser Gln Trp Asn Leu Ser  
85 90 95

Pro Arg Ala Gln Pro Ser Cys Ala Val Thr Val Glu Ser His Thr His  
100 105 110

Ala Ser Pro Arg Met Ala Lys Leu Ala Arg Val Val Gly Leu Val Gln  
115 120 125

Glu Glu Gln Pro Ser Asp Met Thr Asn His Pro Arg Tyr Ser Pro Pro  
130 135 140

Pro Gln Gln Pro Gly Thr Pro Gly Tyr Ala Gln Gly Gln Gln Gln Thr  
145 150 155 160

Tyr Ser Gln Gln Phe Asp Trp Arg Tyr Pro Pro Ser Pro Pro Pro Gln  
165 170 175

Pro Thr Gln Tyr Arg Gln Pro Tyr Glu Ala Leu Gly Gly Thr Arg Pro  
180 185 190

Gly Leu Ile Pro Gly Val Ile Pro Thr Met Thr Pro Pro Pro Gln Met  
195 200 205

110

Val Arg Gln Arg Pro Arg Ala Gly Met Leu Ala Ile Gly Ala Val Thr  
210 215 220

Ile Ala Val Val Ser Ala Gly Ile Gly Gly Ala Ala Ala Ser Leu Val  
225 230 235 240

Gly Phe Asn Arg Ala Pro Ala Gly Pro Ser Gly Gly Pro Val Ala Ala  
245 250 255

Ser Ala Ala Pro Ser Ile Pro Ala Ala Asn Met Pro Pro Gly Ser Val  
260 265 270

Glu Gln Val Ala Ala Lys Val Val Pro Ser Val Val Met Leu Glu Thr  
275 280 285

Asp Leu Gly Arg Gln Ser Glu Glu Gly Ser Gly Ile Ile Leu Ser Ala  
290 295 300

Glu Gly Leu Ile Leu Thr Asn Asn His Val Ile Ala Ala Ala Ala Lys  
305 310 315 320

Pro Pro Leu Gly Ser Pro Pro Pro Lys Thr Thr Val Thr Phe Ser Asp  
325 330 335

Gly Arg Thr Ala Pro Phe Thr Val Val Gly Ala Asp Pro Thr Ser Asp  
340 345 350

Ile Ala Val Val Arg Val Gln Gly Val Ser Gly Leu Thr Pro Ile Ser  
355 360 365

Leu Gly Ser Ser Ser Asp Leu Arg Val Gly Gln Pro Val Leu Ala Ile  
370 375 380

Gly Ser Pro Leu Gly Leu Glu Gly Thr Val Thr Thr Gly Ile Val Ser  
385 390 395 400

Ala Leu Asn Arg Pro Val Ser Thr Thr Gly Glu Ala Gly Asn Gln Asn  
405 410 415

Thr Val Leu Asp Ala Ile Gln Thr Asp Ala Ala Ile Asn Pro Gly Asn  
420 425 430

Ser Gly Gly Ala Leu Val Asn Met Asn Ala Gln Leu Val Gly Val Asn  
435 440 445

111

Ser Ala Ile Ala Thr Leu Gly Ala Asp Ser Ala Asp Ala Gln Ser Gly  
450 455 460

Ser Ile Gly Leu Gly Phe Ala Ile Pro Val Asp Gln Ala Lys Arg Ile  
465 470 475 480

Ala Asp Glu Leu Ile Ser Thr Gly Lys Ala Ser His Ala Ser Leu Gly  
485 490 495

Val Gln Val Thr Asn Asp Lys Asp Thr Pro Gly Ala Lys Ile Val Glu  
500 505 510

Val Val Ala Gly Gly Ala Ala Asn Ala Gly Val Pro Lys Gly Val  
515 520 525

Val Val Thr Lys Val Asp Asp Arg Pro Ile Asn Ser Ala Asp Ala Leu  
530 535 540

Val Ala Ala Val Arg Ser Lys Ala Pro Gly Ala Thr Val Ala Leu Thr  
545 550 555 560

Phe Gln Asp Pro Ser Gly Gly Ser Arg Thr Val Gln Val Thr Leu Gly  
565 570 575

Lys Ala Glu Gln  
580

## (2) INFORMATION FOR SEQ ID NO:76:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 233 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

Met Asn Asp Gly Lys Arg Ala Val Thr Ser Ala Val Leu Val Val Leu  
1 5 10 15

112

Gly Ala Cys Leu Ala Leu Trp Leu Ser Gly Cys Ser Ser Pro Lys Pro  
20 25 30

Asp Ala Glu Glu Gln Gly Val Pro Val Ser Pro Thr Ala Ser Asp Pro  
35 40 45

Ala Leu Leu Ala Glu Ile Arg Gln Ser Leu Asp Ala Thr Lys Gly Leu  
50 55 60

Thr Ser Val His Val Ala Val Arg Thr Thr Gly Lys Val Asp Ser Leu  
65 70 75 80

Leu Gly Ile Thr Ser Ala Asp Val Asp Val Arg Ala Asn Pro Leu Ala  
85 90 95

Ala Lys Gly Val Cys Thr Tyr Asn Asp Glu Gln Gly Val Pro Phe Arg  
100 105 110

Val Gln Gly Asp Asn Ile Ser Val Lys Leu Phe Asp Asp Trp Ser Asn  
115 120 125

Leu Gly Ser Ile Ser Glu Leu Ser Thr Ser Arg Val Leu Asp Pro Ala  
130 135 140

Ala Gly Val Thr Gln Leu Leu Ser Gly Val Thr Asn Leu Gln Ala Gln  
145 150 155 160

Gly Thr Glu Val Ile Asp Gly Ile Ser Thr Thr Lys Ile Thr Gly Thr  
165 170 175

Ile Pro Ala Ser Ser Val Lys Met Leu Asp Pro Gly Ala Lys Ser Ala  
180 185 190

Arg Pro Ala Thr Val Trp Ile Ala Gln Asp Gly Ser His His Leu Val  
195 200 205

Arg Ala Ser Ile Asp Leu Gly Ser Gly Ser Ile Gln Leu Thr Gln Ser  
210 215 220

Lys Trp Asn Glu Pro Val Asn Val Asp  
225 230

(2) INFORMATION FOR SEQ ID NO:77:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 66 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

(2) INFORMATION FOR SEO ID NO:78:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 69 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

Val	Pro	Pro	Ala	Pro	Pro	Leu	Pro	Pro	Leu	Pro	Pro	Ser	Pro	Ile	Ser
1				5						10					15
Cys	Ala	Ser	Pro	Pro	Ser	Pro	Pro	Leu	Pro	Pro	Ala	Pro	Pro	Val	Ala
				20					25						30

Pro Gly Pro Pro Met Pro Pro Leu Asp Pro Trp Pro Pro Ala Pro Pro  
35 40 45

Leu Pro Tyr Ser Thr Pro Pro Gly Ala Pro Leu Pro Pro Ser Pro Pro  
50 55 60

Ser Pro Pro Leu Pro  
65

## (2) INFORMATION FOR SEQ ID NO:79:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 355 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

Met Ser Asn Ser Arg Arg Arg Ser Leu Arg Trp Ser Trp Leu Leu Ser  
1 5 10 15

Val Leu Ala Ala Val Gly Leu Gly Leu Ala Thr Ala Pro Ala Gln Ala  
20 25 30

Ala Pro Pro Ala Leu Ser Gln Asp Arg Phe Ala Asp Phe Pro Ala Leu  
35 40 45

Pro Leu Asp Pro Ser Ala Met Val Ala Gln Val Ala Pro Gln Val Val  
50 55 60

Asn Ile Asn Thr Lys Leu Gly Tyr Asn Asn Ala Val Gly Ala Gly Thr  
65 70 75 80

Gly Ile Val Ile Asp Pro Asn Gly Val Val Leu Thr Asn Asn His Val  
85 90 95

Ile Ala Gly Ala Thr Asp Ile Asn Ala Phe Ser Val Gly Ser Gly Gln  
100 105 110

115

Thr Tyr Gly Val Asp Val Val Gly Tyr Asp Arg Thr Gln Asp Val Ala  
115 120 125

Val Leu Gln Leu Arg Gly Ala Gly Leu Pro Ser Ala Ala Ile Gly  
130 135 140

Gly Gly Val Ala Val Gly Glu Pro Val Val Ala Met Gly Asn Ser Gly  
145 150 155 160

Gly Gln Gly Gly Thr Pro Arg Ala Val Pro Gly Arg Val Val Ala Leu  
165 170 175

Gly Gln Thr Val Gln Ala Ser Asp Ser Leu Thr Gly Ala Glu Glu Thr  
180 185 190

Leu Asn Gly Leu Ile Gln Phe Asp Ala Ala Ile Gln Pro Gly Asp Ser  
195 200 205

Gly Gly Pro Val Val Asn Gly Leu Gly Gln Val Val Gly Met Asn Thr  
210 215 220

Ala Ala Ser Asp Asn Phe Gln Leu Ser Gln Gly Gln Gly Phe Ala  
225 230 235 240

Ile Pro Ile Gly Gln Ala Met Ala Ile Ala Gly Gln Ile Arg Ser Gly  
245 250 255

Gly Gly Ser Pro Thr Val His Ile Gly Pro Thr Ala Phe Leu Gly Leu  
260 265 270

Gly Val Val Asp Asn Asn Gly Asn Gly Ala Arg Val Gln Arg Val Val  
275 280 285

Gly Ser Ala Pro Ala Ala Ser Leu Gly Ile Ser Thr Gly Asp Val Ile  
290 295 300

Thr Ala Val Asp Gly Ala Pro Ile Asn Ser Ala Thr Ala Met Ala Asp  
305 310 315 320

Ala Leu Asn Gly His His Pro Gly Asp Val Ile Ser Val Asn Trp Gln  
325 330 335

Thr Lys Ser Gly Gly Thr Arg Thr Gly Asn Val Thr Leu Ala Glu Gly  
340 345 350

Pro Pro Ala  
355

## (2) INFORMATION FOR SEQ ID NO:80:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 205 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

Ser Pro Lys Pro Asp Ala Glu Glu Gln Gly Val Pro Val Ser Pro Thr  
1 5 10 15

Ala Ser Asp Pro Ala Leu Leu Ala Glu Ile Arg Gln Ser Leu Asp Ala  
20 25 30

Thr Lys Gly Leu Thr Ser Val His Val Ala Val Arg Thr Thr Gly Lys  
35 40 45

Val Asp Ser Leu Leu Gly Ile Thr Ser Ala Asp Val Asp Val Arg Ala  
50 55 60

Asn Pro Leu Ala Ala Lys Gly Val Cys Thr Tyr Asn Asp Glu Gln Gly  
65 70 75 80

Val Pro Phe Arg Val Gln Gly Asp Asn Ile Ser Val Lys Leu Phe Asp  
85 90 95

Asp Trp Ser Asn Leu Gly Ser Ile Ser Glu Leu Ser Thr Ser Arg Val  
100 105 110

Leu Asp Pro Ala Ala Gly Val Thr Gln Leu Leu Ser Gly Val Thr Asn  
115 120 125

Leu Gln Ala Gln Gly Thr Glu Val Ile Asp Gly Ile Ser Thr Thr Lys  
130 135 140

Ile Thr Gly Thr Ile Pro Ala Ser Ser Val Lys Met Leu Asp Pro Gly

117

145                    150                    155                    160  
Ala Lys Ser Ala Arg Pro Ala Thr Val Trp Ile Ala Gln Asp Gly Ser  
165                    170                    175  
His His Leu Val Arg Ala Ser Ile Asp Leu Gly Ser Gly Ser Ile Gln  
180                    185                    190  
Leu Thr Gln Ser Lys Trp Asn Glu Pro Val Asn Val Asp  
195                    200                    205

## (2) INFORMATION FOR SEQ ID NO:81:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 286 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

Gly Asp Ser Phe Trp Ala Ala Ala Asp Gln Met Ala Arg Gly Phe Val  
1                    5                    10                    15  
Leu Gly Ala Thr Ala Gly Arg Thr Thr Leu Thr Gly Glu Gly Leu Gln  
20                    25                    30  
His Ala Asp Gly His Ser Leu Leu Leu Asp Ala Thr Asn Pro Ala Val  
35                    40                    45  
Val Ala Tyr Asp Pro Ala Phe Ala Tyr Glu Ile Tyr Ile Xaa Glu  
50                    55                    60  
Ser Gly Leu Ala Arg Met Cys Gly Glu Asn Pro Glu Asn Ile Phe Phe  
65                    70                    75                    80  
Tyr Ile Thr Val Tyr Asn Glu Pro Tyr Val Gln Pro Pro Glu Pro Glu  
85                    90                    95  
Asn Phe Asp Pro Glu Gly Val Leu Gly Gly Ile Tyr Arg Tyr His Ala  
100                    105                    110

Ala Thr Glu Gln Arg Thr Asn Lys Xaa Gln Ile Leu Ala Ser Gly Val  
115 120 125

Ala Met Pro Ala Ala Leu Arg Ala Ala Gln Met Leu Ala Ala Glu Trp  
130 135 140

Asp Val Ala Ala Asp Val Trp Ser Val Thr Ser Trp Gly Glu Leu Asn  
145 150 155 160

Arg Asp Gly Val Val Ile Glu Thr Glu Lys Leu Arg His Pro Asp Arg  
165 170 175

Pro Ala Gly Val Pro Tyr Val Thr Arg Ala Leu Glu Asn Ala Arg Gly  
180 185 190

Pro Val Ile Ala Val Ser Asp Trp Met Arg Ala Val Pro Glu Gln Ile  
195 200 205

Arg Pro Trp Val Pro Gly Thr Tyr Leu Thr Leu Gly Thr Asp Gly Phe  
210 215 220

Gly Phe Ser Asp Thr Arg Pro Ala Gly Arg Arg Tyr Phe Asn Thr Asp  
225 230 235 240

Ala Glu Ser Gln Val Gly Arg Gly Phe Gly Arg Gly Trp Pro Gly Arg  
245 250 255

Arg Val Asn Ile Asp Pro Phe Gly Ala Gly Arg Gly Pro Pro Ala Gln  
260 265 270

Leu Pro Gly Phe Asp Glu Gly Gly Leu Arg Pro Xaa Lys  
275 280 285

## (2) INFORMATION FOR SEQ ID NO:82:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 173 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

Thr Lys Phe His Ala Leu Met Gln Glu Gln Ile His Asn Glu Phe Thr  
1 5 10 15

Ala Ala Gln Gln Tyr Val Ala Ile Ala Val Tyr Phe Asp Ser Glu Asp  
20 25 30

Leu Pro Gln Leu Ala Lys His Phe Tyr Ser Gln Ala Val Glu Glu Arg  
35 40 45

Asn His Ala Met Met Leu Val Gln His Leu Leu Asp Arg Asp Leu Arg  
50 55 60

Val Glu Ile Pro Gly Val Asp Thr Val Arg Asn Gln Phe Asp Arg Pro  
65 70 75 80

Arg Glu Ala Leu Ala Leu Ala Leu Asp Gln Glu Arg Thr Val Thr Asp  
85 90 95

Gln Val Gly Arg Leu Thr Ala Val Ala Arg Asp Glu Gly Asp Phe Leu  
100 105 110

Gly Glu Gln Phe Met Gln Trp Phe Leu Gln Glu Gln Ile Glu Glu Val  
115 120 125

Ala Leu Met Ala Thr Leu Val Arg Val Ala Asp Arg Ala Gly Ala Asn  
130 135 140

Leu Phe Glu Leu Glu Asn Phe Val Ala Arg Glu Val Asp Val Ala Pro  
145 150 155 160

Ala Ala Ser Gly Ala Pro His Ala Ala Gly Gly Arg Leu  
165 170

## (2) INFORMATION FOR SEQ ID NO:83:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 107 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

120

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

Arg Ala Asp Glu Arg Lys Asn Thr Thr Met Lys Met Val Lys Ser Ile  
1                   5                   10                   15

Ala Ala Gly Leu Thr Ala Ala Ala Ile Gly Ala Ala Ala Ala Gly  
20                  25                  30

Val Thr Ser Ile Met Ala Gly Gly Pro Val Val Tyr Gln Met Gln Pro  
35                  40                  45

Val Val Phe Gly Ala Pro Leu Pro Leu Asp Pro Xaa Ser Ala Pro Xaa  
50                  55                  60

Val Pro Thr Ala Ala Gln Trp Thr Xaa Leu Leu Asn Xaa Leu Xaa Asp  
65                  70                  75                  80

Pro Asn Val Ser Phe Xaa Asn Lys Gly Ser Leu Val Glu Gly Gly Ile  
85                  90                  95

Gly Gly Xaa Glu Gly Xaa Xaa Arg Arg Xaa Gln  
100                105

## (2) INFORMATION FOR SEQ ID NO:84:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 125 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

Val Leu Ser Val Pro Val Gly Asp Gly Phe Trp Xaa Arg Val Val Asn  
1                  5                  10                  15

Pro Leu Gly Gln Pro Ile Asp Gly Arg Gly Asp Val Asp Ser Asp Thr  
20                25                  30

121

Arg Arg Ala Leu Glu Leu Gln Ala Pro Ser Val Val Xaa Arg Gln Gly  
35 40 45

Val Lys Glu Pro Leu Xaa Thr Gly Ile Lys Ala Ile Asp Ala Met Thr  
50 55 60

Pro Ile Gly Arg Gly Gln Arg Gln Leu Ile Ile Gly Asp Arg Lys Thr  
65 70 75 80

Gly Lys Asn Arg Arg Leu Cys Arg Thr Pro Ser Ser Asn Gln Arg Glu  
85 90 95

Glu Leu Gly Val Arg Trp Ile Pro Arg Ser Arg Cys Ala Cys Val Tyr  
100 105 110

Val Gly His Arg Ala Arg Arg Gly Thr Tyr His Arg Arg  
115 120 125

## (2) INFORMATION FOR SEQ ID NO:85:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 117 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

Cys Asp Ala Val Met Gly Phe Leu Gly Gly Ala Gly Pro Leu Ala Val  
1 5 10 15

Val Asp Gln Gln Leu Val Thr Arg Val Pro Gln Gly Trp Ser Phe Ala  
20 25 30

Gln Ala Ala Ala Val Pro Val Val Phe Leu Thr Ala Trp Tyr Gly Leu  
35 40 45

Ala Asp Leu Ala Glu Ile Lys Ala Gly Glu Ser Val Leu Ile His Ala  
50 55 60

Gly Thr Gly Gly Val Gly Met Ala Ala Val Gln Leu Ala Arg Gln Trp

122

65	70	75	80
Gly Val Glu Val Phe Val Thr Ala Ser Arg Gly Lys Trp Asp Thr Leu			
85	90	95	
Arg Ala Xaa Xaa Phe Asp Asp Xaa Pro Tyr Arg Xaa Phe Pro His Xaa			
100	105	110	
Arg Ser Ser Xaa Gly			
115			

## (2) INFORMATION FOR SEQ ID NO:86:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 103 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

Met Tyr Arg Phe Ala Cys Arg Thr Leu Met Leu Ala Ala Cys Ile Leu			
1	5	10	15
Ala Thr Gly Val Ala Gly Leu Gly Val Gly Ala Gln Ser Ala Ala Gln			
20	25	30	
Thr Ala Pro Val Pro Asp Tyr Tyr Trp Cys Pro Gly Gln Pro Phe Asp			
35	40	45	
Pro Ala Trp Gly Pro Asn Trp Asp Pro Tyr Thr Cys His Asp Asp Phe			
50	55	60	
His Arg Asp Ser Asp Gly Pro Asp His Ser Arg Asp Tyr Pro Gly Pro			
65	70	75	80
Ile Leu Glu Gly Pro Val Leu Asp Asp Pro Gly Ala Ala Pro Pro Pro			
85	90	95	
Pro Ala Ala Gly Gly Gly Ala			
100			

## (2) INFORMATION FOR SEQ ID NO:87:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 88 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

Val Gln Cys Arg Val Trp Leu Glu Ile Gln Trp Arg Gly Met Leu Gly  
1               5                           10                           15

Ala Asp Gln Ala Arg Ala Gly Gly Pro Ala Arg Ile Trp Arg Glu His  
20               25                           30

Ser Met Ala Ala Met Lys Pro Arg Thr Gly Asp Gly Pro Leu Glu Ala  
35               40                           45

Thr Lys Glu Gly Arg Gly Ile Val Met Arg Val Pro Leu Glu Gly Gly  
50               55                           60

Gly Arg Leu Val Val Glu Leu Thr Pro Asp Glu Ala Ala Leu Gly  
65               70                           75                           80

Asp Glu Leu Lys Gly Val Thr Ser  
85

## (2) INFORMATION FOR SEQ ID NO:88:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 95 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

Thr Asp Ala Ala Thr Leu Ala Gln Glu Ala Gly Asn Phe Glu Arg Ile  
1               5               10               15

Ser Gly Asp Leu Lys Thr Gln Ile Asp Gln Val Glu Ser Thr Ala Gly  
20               25               30

Ser Leu Gln Gly Gln Trp Arg Gly Ala Ala Gly Thr Ala Ala Gln Ala  
35               40               45

Ala Val Val Arg Phe Gln Glu Ala Ala Asn Lys Gln Lys Gln Glu Leu  
50               55               60

Asp Glu Ile Ser Thr Asn Ile Arg Gln Ala Gly Val Gln Tyr Ser Arg  
65               70               75               80

Ala Asp Glu Glu Gln Gln Ala Leu Ser Ser Gln Met Gly Phe  
85               90               95

## (2) INFORMATION FOR SEQ ID NO:89:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 166 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

Met Thr Gln Ser Gln Thr Val Thr Val Asp Gln Gln Glu Ile Leu Asn  
1               5               10               15

Arg Ala Asn Glu Val Glu Ala Pro Met Ala Asp Pro Pro Thr Asp Val  
20               25               30

Pro Ile Thr Pro Cys Glu Leu Thr Xaa Xaa Lys Asn Ala Ala Gln Gln  
35               40               45

Xaa Val Leu Ser Ala Asp Asn Met Arg Glu Tyr Leu Ala Ala Gly Ala  
50               55               60

125

Lys Glu Arg Gln Arg Leu Ala Thr Ser Leu Arg Asn Ala Ala Lys Xaa  
65                   70                   75                   80

Tyr Gly Glu Val Asp Glu Glu Ala Ala Thr Ala Leu Asp Asn Asp Gly  
85                   90                   95

Glu Gly Thr Val Gln Ala Glu Ser Ala Gly Ala Val Gly Gly Asp Ser  
100               105               110

Ser Ala Glu Leu Thr Asp Thr Pro Arg Val Ala Thr Ala Gly Glu Pro  
115               120               125

Asn Phe Met Asp Leu Lys Glu Ala Ala Arg Lys Leu Glu Thr Gly Asp  
130               135               140

Gln Gly Ala Ser Leu Ala His Xaa Gly Asp Gly Trp Asn Thr Xaa Thr  
145               150               155               160

Leu Thr Leu Gln Gly Asp  
165

## (2) INFORMATION FOR SEQ ID NO:90:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 5 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

Arg Ala Glu Arg Met  
1               5

## (2) INFORMATION FOR SEQ ID NO:91:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 263 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single

126

## (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

Val Ala Trp Met Ser Val Thr Ala Gly Gln Ala Glu Leu Thr Ala Ala  
1 5 10 15

Gln Val Arg Val Ala Ala Ala Tyr Glu Thr Ala Tyr Gly Leu Thr  
20 25 30

Val Pro Pro Pro Val Ile Ala Glu Asn Arg Ala Glu Leu Met Ile Leu  
35 40 45

Ile Ala Thr Asn Leu Leu Gly Gln Asn Thr Pro Ala Ile Ala Val Asn  
50 55 60

Glu Ala Glu Tyr Gly Glu Met Trp Ala Gln Asp Ala Ala Ala Met Phe  
65 70 75 80

Gly Tyr Ala Ala Ala Thr Ala Thr Ala Thr Leu Leu Pro Phe  
85 90 95

Glu Glu Ala Pro Glu Met Thr Ser Ala Gly Gly Leu Leu Glu Gln Ala  
100 105 110

Ala Ala Val Glu Glu Ala Ser Asp Thr Ala Ala Ala Asn Gln Leu Met  
115 120 125

Asn Asn Val Pro Gln Ala Leu Lys Gln Leu Ala Gln Pro Thr Gln Gly  
130 135 140

Thr Thr Pro Ser Ser Lys Leu Gly Gly Leu Trp Lys Thr Val Ser Pro  
145 150 155 160

His Arg Ser Pro Ile Ser Asn Met Val Ser Met Ala Asn Asn His Met  
165 170 175

Ser Met Thr Asn Ser Gly Val Ser Met Thr Asn Thr Leu Ser Ser Met  
180 185 190

Leu Lys Gly Phe Ala Pro Ala Ala Ala Gln Ala Val Gln Thr Ala

127

195

200

205

Ala Gln Asn Gly Val Arg Ala Met Ser Ser Leu Gly Ser Ser Leu Gly  
210 215 220

Ser Ser Gly Leu Gly Gly Val Ala Ala Asn Leu Gly Arg Ala Ala  
225 230 235 240

Ser Val Arg Tyr Gly His Arg Asp Gly Gly Lys Tyr Ala Xaa Ser Gly  
245 250 255

Arg Arg Asn Gly Gly Pro Ala  
260

(2) INFORMATION FOR SEQ ID NO:92:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 303 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:

Met Thr Tyr Ser Pro Gly Asn Pro Gly Tyr Pro Gln Ala Gln Pro Ala  
1 5 10 15

Gly Ser Tyr Gly Gly Val Thr Pro Ser Phe Ala His Ala Asp Glu Gly  
20 25 30

Ala Ser Lys Leu Pro Met Tyr Leu Asn Ile Ala Val Ala Val Leu Gly  
35 40 45

Leu Ala Ala Tyr Phe Ala Ser Phe Gly Pro Met Phe Thr Leu Ser Thr  
50 55 60

Glu Leu Gly Gly Asp Gly Ala Val Ser Gly Asp Thr Gly Leu Pro  
65 70 75 80

Val Gly Val Ala Leu Leu Ala Ala Leu Leu Ala Gly Val Val Leu Val  
85 90 95

128

Pro	Lys	Ala	Lys	Ser	His	Val	Thr	Val	Val	Ala	Val	Leu	Gly	Val	Leu
						100			105						110
Gly	Val	Phe	Leu	Met	Val	Ser	Ala	Thr	Phe	Asn	Lys	Pro	Ser	Ala	Tyr
						115			120						125
Ser	Thr	Gly	Trp	Ala	Leu	Trp	Val	Val	Leu	Ala	Phe	Ile	Val	Phe	Gln
						130			135						140
Ala	Val	Ala	Ala	Val	Leu	Ala	Leu	Leu	Val	Glu	Thr	Gly	Ala	Ile	Thr
						145			150						160
Ala	Pro	Ala	Pro	Arg	Pro	Lys	Phe	Asp	Pro	Tyr	Gly	Gln	Tyr	Gly	Arg
						165			170						175
Tyr	Gly	Gln	Tyr	Gly	Gln	Tyr	Gly	Val	Gln	Pro	Gly	Gly	Tyr	Tyr	Gly
						180			185						190
Gln	Gln	Gly	Ala	Gln	Gln	Ala	Ala	Gly	Leu	Gln	Ser	Pro	Gly	Pro	Gln
						195			200						205
Gln	Ser	Pro	Gln	Pro	Pro	Gly	Tyr	Gly	Ser	Gln	Tyr	Gly	Gly	Tyr	Ser
						210			215						220
Ser	Ser	Pro	Ser	Gln	Ser	Gly	Ser	Gly	Tyr	Thr	Ala	Gln	Pro	Pro	Ala
						225			230						240
Gln	Pro	Pro	Ala	Gln	Ser	Gly	Ser	Gln	Gln	Ser	His	Gln	Gly	Pro	Ser
						245			250						255
Thr	Pro	Pro	Thr	Gly	Phe	Pro	Ser	Phe	Ser	Pro	Pro	Pro	Pro	Val	Ser
						260			265						270
Ala	Gly	Thr	Gly	Ser	Gln	Ala	Gly	Ser	Ala	Pro	Val	Asn	Tyr	Ser	Asn
						275			280						285
Pro	Ser	Gly	Gly	Glu	Gln	Ser	Ser	Ser	Pro	Gly	Gly	Ala	Pro	Val	
						290			295						300

## (2) INFORMATION FOR SEQ ID NO:93:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 28 amino acids
  - (B) TYPE: amino acid

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:93:

Gly	Cys	Gly	Glu	Thr	Asp	Ala	Ala	Thr	Leu	Ala	Gln	Glu	Ala	Gly	Asn
1				5					10					15	
Phe	Glu	Arg	Ile	Ser	Gly	Asp	Leu	Lys	Thr	Gln	Ile				
				20				25							

(2) INFORMATION FOR SEQ ID NO:94:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 16 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

Asp	Gln	Val	Glu	Ser	Thr	Ala	Gly	Ser	Leu	Gln	Gly	Gln	Trp	Arg	Gly
1				5					10				15		

(2) INFORMATION FOR SEQ ID NO:95:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 27 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

130

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

Gly Cys Gly Ser Thr Ala Gly Ser Leu Gln Gly Gln Trp Arg Gly Ala  
1               5               10                           15  
  
Ala Gly Thr Ala Ala Gln Ala Ala Val Val Arg  
20                           25

## (2) INFORMATION FOR SEQ ID NO:96:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 27 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:

Gly Cys Gly Gly Thr Ala Ala Gln Ala Ala Val Val Arg Phe Gln Glu  
1               5               10                           15  
  
Ala Ala Asn Lys Gln Lys Gln Glu Leu Asp Glu  
20                           25

## (2) INFORMATION FOR SEQ ID NO:97:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 27 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:97:

Gly Cys Gly Ala Asn Lys Gln Lys Gln Glu Leu Asp Glu Ile Ser Thr  
1               5               10                           15

Asn Ile Arg Gln Ala Gly Val Gln Tyr Ser Arg  
20   25

## (2) INFORMATION FOR SEQ ID NO:98:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 28 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:98:

Gly Cys Gly Ile Arg Gln Ala Gly Val Gln Tyr Ser Arg Ala Asp Glu  
1                         5                                   10                                   15  
  
Glu Gln Gln Gln Ala Leu Ser Ser Gln Met Gly Phe  
20                         25

## (2) INFORMATION FOR SEQ ID NO:99:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 507 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

ATGAAGATGG TGAAATCGAT CGCCGCAGGT CTGACCGCCG CGGCTGCAAT CGGCGCCGCT       60  
CGGGCCGGTG TGACTTCGAT CATGGCTGGC GGCCCCGTCG TATACCAGAT GCAGCCGGTC       120  
GTCTTCGGCG CGCCACTGCC GTTGGACCCG GCATCCGCCCT CTGACGTCCC GACCGCCGCC       180  
CAGTTGACCA GCCTGCTCAA CAGCCTCGCC GATCCAACG TGTCGTTGC GAACAAGGGC       240

132

AGTCTGGTCG AGGGCGGCAT CGGGGGCACC GAGGCGCGCA TCGCCGACCA CAAGCTGAAG	300
AAGGCCGCCG AGCACGGGGA TCTGCCGCTG TCGTTCAGCG TGACGAACAT CCAGCCGGCG	360
GCCGCCGGTT CGGCCACCGC CGACGTTCC GTCTCGGGTC CGAAGCTCTC GTCGCCGGTC	420
ACGCAGAACG TCACGTTCGT GAATCAAGGC GGCTGGATGC TGTCACGCGC ATCGGCGATG	480
GAGTTGCTGC AGGCCGCAGG GAACTGA	507

## (2) INFORMATION FOR SEQ ID NO:100:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 168 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

Met Lys Met Val Lys Ser Ile Ala Ala Gly Leu Thr Ala Ala Ala Ala			
1	5	10	15
Ile Gly Ala Ala Ala Gly Val Thr Ser Ile Met Ala Gly Gly Pro			
20	25	30	
Val Val Tyr Gln Met Gln Pro Val Val Phe Gly Ala Pro Leu Pro Leu			
35	40	45	
Asp Pro Ala Ser Ala Pro Asp Val Pro Thr Ala Ala Gln Leu Thr Ser			
50	55	60	
Leu Leu Asn Ser Leu Ala Asp Pro Asn Val Ser Phe Ala Asn Lys Gly			
65	70	75	80
Ser Leu Val Glu Gly Gly Ile Gly Gly Thr Glu Ala Arg Ile Ala Asp			
85	90	95	
His Lys Leu Lys Lys Ala Ala Glu His Gly Asp Leu Pro Leu Ser Phe			
100	105	110	

Ser Val Thr Asn Ile Gln Pro Ala Ala Ala Gly Ser Ala Thr Ala Asp  
 115                    120                    125  
 Val Ser Val Ser Gly Pro Lys Leu Ser Ser Pro Val Thr Gln Asn Val  
 130                    135                    140  
 Thr Phe Val Asn Gln Gly Gly Trp Met Leu Ser Arg Ala Ser Ala Met  
 145                    150                    155                    160  
 Glu Leu Leu Gln Ala Ala Gly Asn  
 165

## (2) INFORMATION FOR SEQ ID NO:101:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 500 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:101:

CGTGGCAATG TCGTTGACCG TCGGGGCCGG GGTCGCCTCC GCAGATCCCG TGGACGCGGT	60
CATTAACACC ACCTGCAATT ACGGGCAGGT AGTAGCTGCG CTCAACGCGA CGGATCCGGG	120
GGCTGCCGCA CAGTTAACG CCTCACCGGT GGCGCAGTCC TATTTGCGCA ATTTCTCGC	180
CGCACCGCCA CCTCAGCGCG CTGCCATGGC CGCGCAATTG CAAGCTGTGC CGGGGGCGGC	240
ACAGTACATC GGCTTGTCG AGTCGGTTGC CGGCTCCTGC AACAACTATT AAGCCCATGC	300
GGGCCCCATC CCGCGACCCG GCATCGTCGC CGGGGCTAGG CCAGATTGCC CCGCTCCTCA	360
ACGGGCGCA TCCCGCGACC CGGCATCGTC GCCGGGGCTA GGCCAGATTG CCCCCTCCT	420
CAACGGGCCG CATCTCGTGC CGAATTCTG CAGCCCGGGG GATCCACTAG TTCTAGAGCG	480
GCCGCCACCG CGGTGGAGCT	500

## (2) INFORMATION FOR SEQ ID NO:102:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 96 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:102:

Val	Ala	Met	Ser	Leu	Thr	Val	Gly	Ala	Gly	Val	Ala	Ser	Ala	Asp	Pro
1				5				10						15	
Val	Asp	Ala	Val	Ile	Asn	Thr	Thr	Cys	Asn	Tyr	Gly	Gln	Val	Val	Ala
	20				25							30			
Ala	Leu	Asn	Ala	Thr	Asp	Pro	Gly	Ala	Ala	Ala	Gln	Phe	Asn	Ala	Ser
	35				40						45				
Pro	Val	Ala	Gln	Ser	Tyr	Leu	Arg	Asn	Phe	Leu	Ala	Ala	Pro	Pro	Pro
	50				55				60						
Gln	Arg	Ala	Ala	Met	Ala	Ala	Gln	Leu	Gln	Ala	Val	Pro	Gly	Ala	Ala
	65				70				75		80				
Gln	Tyr	Ile	Gly	Leu	Val	Glu	Ser	Val	Ala	Gly	Ser	Cys	Asn	Asn	Tyr
	85					90						95			

## (2) INFORMATION FOR SEQ ID NO:103:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 154 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:103:

135

ATGACAGAGC AGCAGTGGAA TTTCGCGGGT ATCGAGGCCG CGGCAAGCGC AATCCAGGGA	60
AATGTCACGT CCATTCAATT CCTCCTTGAC GAGGGGAAGC AGTCCCTGAC CAAGCTCGCA	120
GCGGCCTGGG GCGGTAGCGG TTCCGAAGCG TACC	154

## (2) INFORMATION FOR SEQ ID NO:104:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 51 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:104:

Met Thr Glu Gln Gln Trp Asn Phe Ala Gly Ile Glu Ala Ala Ala Ser			
1	5	10	15
Ala Ile Gln Gly Asn Val Thr Ser Ile His Ser Leu Leu Asp Glu Gly			
20	25	30	
Lys Gln Ser Leu Thr Lys Leu Ala Ala Ala Trp Gly Gly Ser Gly Ser			
35	40	45	
Glu Ala Tyr			
50			

## (2) INFORMATION FOR SEQ ID NO:105:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 282 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:105:

CGGTGCGCA CTTCCAGGTG ACTATGAAAG TCGGCTTCG NCTGGAGGAT TCCTAACCT	60
TCAAGCGCG CCGATAACTG AGGTGCATCA TTAAGCGACT TTTCCAGAAC ATCCTGACGC	120
GCTCGAAACG CGGCACAGCC GACGGTGGCT CCGNCAGGC GCTGNCTCCA AAATCCCTGA	180
GACAATTCGN CGGGGGCGCC TACAAGGAAG TCGGTGCTGA ATTGNCGNG TATCTGGTCG	240
ACCTGTGTGG TCTGNAGCCG GACGAAGCGG TGCTCGACGT CG	282

## (2) INFORMATION FOR SEQ ID NO:106:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1565 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:106:

GTATGCGGCC ACTGAAGTCG CCAATGCGC GGCGGCCAGC TAAGCCAGGA ACAGTCGGCA	60
CGAGAAACCA CGAGAAATAG GGACACGTA TGGTGGATTT CGGGCGTTA CCACCGGAGA	120
TCAACTCCGC GAGGATGTAC GCCGGCCCCG GTTCGGCTC GCTGGTGGCC GCGGCTCAGA	180
TGTGGGACAG CGTGGCGAGT GACCTGTTT CGGCCGCGTC GGCCTTCAG TCGGTGGTCT	240
GGGGTCTGAC GGTGGGGTCG TGGATAGGTT CGTCGGCGGG TCTGATGGTG GCGGCGGCCT	300
CGCCGTATGT GGCGTGGATG AGCGTCACCG CGGGGCAGGC CGAGCTGACC GCCGCCCAGG	360
TCCGGGTTGC TGCAGCGGCC TACGAGACGG CGTATGGCT GACGGTGCCC CCGCCGGTGA	420
TCGCCGAGAA CCGTGCTGAA CTGATGATTG TGATAGCGAC CAACCTCTTG GGGCAAAACA	480
CCCCGGCGAT CGCGGTCAAC GAGGCCGAAT ACGGCGAGAT GTGGGCCAA GACGCCGCG	540
CGATGTTGG CTACGCCGCG GCGACGGCGA CGGCGACGGC GACGTTGCTG CCGTTCGAGG	600

AGGCGCCGGA GATGACCAGC GCGGGTGGC TCCTCGAGCA GGCGCCGCG GTCGAGGAGG	660
CCTCCGACAC CGCCGCGGCG AACCAGTTGA TGAACAATGT GCCCCAGGCG CTGCAACAGC	720
TGGCCCAGCC CACGCAGGGC ACCACGCCCTT CTTCCAAGCT GGGTGGCCTG TGGAAAGACGG	780
TCTCGCCGCA TCGGTCGCG ATCAGCAACA TGGTGTCAAT GGCCAACAAAC CACATGTCAA	840
TGACCAACTC GGGTGTGTCA ATGACCAACA CCTTGAGCTC GATGTTGAAG GGCTTGCTC	900
CGGCGGCGGC CGCCCAGGCC GTGCAAACCG CGGCGCAAAA CGGGGTCCGG GCGATGAGCT	960
CGCTGGGCAG CTCGCTGGGT TCTTCGGGTC TGGGCGGTGG GGTGGCCGCC AACTTGGTC	1020
GGGCGGCCTC GGTCGGTTCG TTGTCGGTGC CGCAGGCCTG GGCCGCGGCC AACCAAGGCAG	1080
TCACCCCCGGC GGCGCGGGCG CTGCCGCTGA CCAGCCTGAC CAGCGCCGCG GAAAGAGGGC	1140
CGGGGCAGAT GCTGGGCGGG CTGCCGGTGG GGCAGATGGG CGCCAGGGCC GGTGGTGGC	1200
TCAGTGGTGT GCTGCGTGT CCGCCGCGAC CCTATGTGAT GCCGCATTCT CCGGCGGCCG	1260
GCTAGGAGAG GGGCGCAGA CTGTCGTTAT TTGACCAGTG ATCGGCGGTC TCGGTGTTTC	1320
CGCGGCCGGC TATGACAACA GTCAATGTGC ATGACAAGTT ACAGGTATTA GGTCCAGGTT	1380
CAACAAGGAG ACAGGCAACA TGGCCTCACG TTTTATGACG GATCCGCACG CGATGCGGGA	1440
CATGGCGGGC CGTTTGAAAG TGACGCCCA GACGGTGGAG GACGAGGCTC GCCGGATGTG	1500
GGCGTCCGCG CAAAACATT CCGGTGCGGG CTGGAGTGGC ATGGCCGAGG CGACCTCGCT	1560
AGACA	1565

## (2) INFORMATION FOR SEQ ID NO:107:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 391 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:107:

Met Val Asp Phe Gly Ala Leu Pro Pro Glu Ile Asn Ser Ala Arg Met  
1 5 10 15

Tyr Ala Gly Pro Gly Ser Ala Ser Leu Val Ala Ala Ala Gln Met Trp  
20 25 30

Asp Ser Val Ala Ser Asp Leu Phe Ser Ala Ala Ser Ala Phe Gln Ser  
35 40 45

Val Val Trp Gly Leu Thr Val Gly Ser Trp Ile Gly Ser Ser Ala Gly  
50 55 60

Leu Met Val Ala Ala Ala Ser Pro Tyr Val Ala Trp Met Ser Val Thr  
65 70 75 80

Ala Gly Gln Ala Glu Leu Thr Ala Ala Gln Val Arg Val Ala Ala Ala  
85 90 95

Ala Tyr Glu Thr Ala Tyr Gly Leu Thr Val Pro Pro Pro Val Ile Ala  
100 105 110

Glu Asn Arg Ala Glu Leu Met Ile Leu Ile Ala Thr Asn Leu Leu Gly  
115 120 125

Gln Asn Thr Pro Ala Ile Ala Val Asn Glu Ala Glu Tyr Gly Glu Met  
130 135 140

Trp Ala Gln Asp Ala Ala Ala Met Phe Gly Tyr Ala Ala Ala Thr Ala  
145 150 155 160

Thr Ala Thr Ala Thr Leu Leu Pro Phe Glu Glu Ala Pro Glu Met Thr  
165 170 175

Ser Ala Gly Gly Leu Leu Glu Gln Ala Ala Ala Val Glu Glu Ala Ser  
180 185 190

Asp Thr Ala Ala Ala Asn Gln Leu Met Asn Asn Val Pro Gln Ala Leu  
195 200 205

Gln Gln Leu Ala Gln Pro Thr Gln Gly Thr Thr Pro Ser Ser Lys Leu  
210 215 220

139

Gly Gly Leu Trp Lys Thr Val Ser Pro His Arg Ser Pro Ile Ser Asn  
225 230 235 240

Met Val Ser Met Ala Asn Asn His Met Ser Met Thr Asn Ser Gly Val  
245 250 255

Ser Met Thr Asn Thr Leu Ser Ser Met Leu Lys Gly Phe Ala Pro Ala  
260 265 270

Ala Ala Ala Gln Ala Val Gln Thr Ala Ala Gln Asn Gly Val Arg Ala  
275 280 285

Met Ser Ser Leu Gly Ser Ser Leu Gly Ser Ser Gly Leu Gly Gly  
290 295 300

Val Ala Ala Asn Leu Gly Arg Ala Ala Ser Val Gly Ser Leu Ser Val  
305 310 315 320

Pro Gln Ala Trp Ala Ala Asn Gln Ala Val Thr Pro Ala Ala Arg  
325 330 335

Ala Leu Pro Leu Thr Ser Leu Thr Ser Ala Ala Glu Arg Gly Pro Gly  
340 345 350

Gln Met Leu Gly Gly Leu Pro Val Gly Gln Met Gly Ala Arg Ala Gly  
355 360 365

Gly Gly Leu Ser Gly Val Leu Arg Val Pro Pro Arg Pro Tyr Val Met  
370 375 380

Pro His Ser Pro Ala Ala Gly  
385 390

## (2) INFORMATION FOR SEQ ID NO:108:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 259 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

140

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:108:

ACCAACACCT TGCACTCNAT GTTGAAGGGC TTAGCTCCGG CGGCGGCTCA GGCGTGGAA	60
ACCGCGGCCG AAAACGGGGT CTGGGCAATG AGCTCGCTGG GCAGCCAGCT GGGTCGTGCG	120
CTGGGTTCTT CGGGTCTGGG CGCTGGGTG GCCGCCAACT TGGGTCGGGC GGCGCTCGTC	180
GGTCGTTGT CGGTGCCGCC AGCATGGGCC GCGGCCAACC AGGCGGTAC CCCGGCGCG	240
CGGGCGCTGC CGCTGACCA	259

## (2) INFORMATION FOR SEQ ID NO:109:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 86 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:109:

Thr Asn Thr Leu His Ser Met Leu Lys Gly Leu Ala Pro Ala Ala Ala			
1	5	10	15
Gln Ala Val Glu Thr Ala Ala Glu Asn Gly Val Trp Ala Met Ser Ser			
20	25	30	
Leu Gly Ser Gln Leu Gly Ser Ser Leu Gly Ser Ser Gly Leu Gly Ala			
35	40	45	
Gly Val Ala Ala Asn Leu Gly Arg Ala Ala Ser Val Gly Ser Leu Ser			
50	55	60	
Val Pro Pro Ala Trp Ala Ala Ala Asn Gln Ala Val Thr Pro Ala Ala			
65	70	75	80
Arg Ala Leu Pro Leu Thr			
85			

## (2) INFORMATION FOR SEQ ID NO:110:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1109 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:110:

TACTTGAGAG AATTGACCT GTTGCCGACG TTGTTGCTG TCCATCATTG GTGCTAGTTA	60
TGGCCGAGCG GAAGGATTAT CGAAAGTGGTG GACTTCGGGG CGTTACCACC GGAGATCAAC	120
TCCGCGAGGA TGTACGCCGG CCCGGGTTCG GCCTCGCTGG TGGCCGCCGC GAAGATGTGG	180
GACAGCGTGG CGAGTGACCT GTTTTGGCC GCGTCGGCGT TTCAGTCGGT GGTCTGGGT	240
CTGACGACGG GATCGTGGAT AGGTTCGTCG GCGGGTCTGA TGGTGGCGGC GGCCTCGCCG	300
TATGTGGCGT GGATGAGCGT CACCGCGGGG CAGGCCGAGC TGACCGCCGC CCAGGTCCGG	360
GTTGCTGCGG CGGCCTACGA GACGGCGTAT GGGCTGACGG TGCCCCCGCC GGTGATGCC	420
GAGAACCGTG CTGAAC TGAT GATTCTGATA GCGACCAACC TCTTGGGGCA AAACACCCCG	480
GCGATCGCGG TCAACGAGGC CGAACATCGGG GAGATGTGGG CCCAAGACGC CGCCGCGATG	540
TTTGGCTACG CCGCCACGGC GGCGACGGCG ACCGAGGC GT TGCTGCCGTT CGAGGACGCC	600
CCACTGATCA CCAACCCCGG CGGGCTCCTT GAGCAGGCCG TCGCGGTGAA GGAGGCCATC	660
GACACCGCCG CGCGAACCA GTTGATGAAAC AATGTGCCCC AAGCGCTGCA ACAACTGCC	720
CAGCCCACGA AAAGCATCTG GCCGTTGAC CAACTGAGTG AACTCTGGAA AGCCATCTG	780
CCGCATCTGT CGCCGCTAG CAACATCGTG TCGATGCTCA ACAACCACGT GTCGATGACC	840
AACTCGGGTG TGTCAATGGC CAGCACCTTG CACTCAATGT TGAAGGGCTT TGCTCCGGCG	900
GC GGCTCAGG CCGTGGAAAC CGCGGCGCAA AACGGGGTCC AGGCGATGAG CTCGCTGGC	960

AGCCAGCTGG GTTCGTCGCT GGGTTCTTCG GGTCTGGCG CTGGGGTGGC CGCCAACTTG 1020  
GGTCGGGCGG CCTCGGTGCG TTCTGGTGTG GTGCCGCAGG CCTGGGCCGC GGCCAACCAG 1080  
GCGGTACCCC CGGCAGCGCG GGCGCTGCC 1109

(2) INFORMATION FOR SEQ ID NO:111:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 341 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:111:

Val Val Asp Phe Gly Ala Leu Pro Pro Glu Ile Asn Ser Ala Arg Met  
1 5 10 15

Tyr Ala Gly Pro Gly Ser Ala Ser Leu Val Ala Ala Ala Lys Met Trp  
20 25 30

Asp Ser Val Ala Ser Asp Leu Phe Ser Ala Ala Ser Ala Phe Gln Ser  
35 40 45

Val Val Trp Gly Leu Thr Thr Gly Ser Trp Ile Gly Ser Ser Ala Gly  
50 55 60

Leu Met Val Ala Ala Ala Ser Pro Tyr Val Ala Trp Met Ser Val Thr  
65                  70                  75                  80

Ala Gly Gln Ala Glu Leu Thr Ala Ala Gln Val Arg Val Ala Ala Ala  
85 90 95

Ala Tyr Glu Thr Ala Tyr Gly Leu Thr Val Pro Pro Pro Val Ile Ala  
100 105 110

Glu Asn Arg Ala Glu Leu Met Ile Leu Ile Ala Thr Asn Leu Leu Gly  
115 120 125

Gln Asn Thr Pro Ala Ile Ala Val Asn Glu Ala Glu Tyr Gly Glu Met  
 130 135 140  
 Trp Ala Gln Asp Ala Ala Ala Met Phe Gly Tyr Ala Ala Thr Ala Ala  
 145 150 155 160  
 Thr Ala Thr Glu Ala Leu Leu Pro Phe Glu Asp Ala Pro Leu Ile Thr  
 165 170 175  
 Asn Pro Gly Gly Leu Leu Glu Gln Ala Val Ala Val Glu Glu Ala Ile  
 180 185 190  
 Asp Thr Ala Ala Ala Asn Gln Leu Met Asn Asn Val Pro Gln Ala Leu  
 195 200 205  
 Gln Gln Leu Ala Gln Pro Thr Lys Ser Ile Trp Pro Phe Asp Gln Leu  
 210 215 220  
 Ser Glu Leu Trp Lys Ala Ile Ser Pro His Leu Ser Pro Leu Ser Asn  
 225 230 235 240  
 Ile Val Ser Met Leu Asn Asn His Val Ser Met Thr Asn Ser Gly Val  
 245 250 255  
 Ser Met Ala Ser Thr Leu His Ser Met Leu Lys Gly Phe Ala Pro Ala  
 260 265 270  
 Ala Ala Gln Ala Val Glu Thr Ala Ala Gln Asn Gly Val Gln Ala Met  
 275 280 285  
 Ser Ser Leu Gly Ser Gln Leu Gly Ser Ser Leu Gly Ser Ser Gly Leu  
 290 295 300  
 Gly Ala Gly Val Ala Ala Asn Leu Gly Arg Ala Ala Ser Val Gly Ser  
 305 310 315 320  
 Leu Ser Val Pro Gln Ala Trp Ala Ala Ala Asn Gln Ala Val Thr Pro  
 325 330 335  
 Ala Ala Arg Ala Leu  
 340

## (2) INFORMATION FOR SEQ ID NO:112:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1256 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:112:

CATCGGAGGG AGTGATCAC	60
ATGCTGTGGC ACGCAATGCC ACCGGAGNTA AATACCGCAC	
GGCTGATGGC CGGCGCGGGT CCGGCTCCAA TGCTTGC	120
GGC CGGGGA TGGCAGACGC	
TTTCGGCGGC TCTGGACGCT CAGGCCGTCG AGTTGACCGC GCGCCTGAAC TCTCTGGGAG	180
AAGCCTGGAC TGGAGGTGGC AGCGACAAGG CGCTTGC	240
GGC TGCGC TGCAACGCG ATGGTGGTCT	
GGCTACAAAC CGCGTCAACA CAGGCCAAGA CCCGTGCGAT GCAGGCGACG GCGCAAGCCG	300
CGGCATACAC CCAGGCCATG GCCACGACGC CGTCGCTGCC GGAGATGCC	360
GCC ACCACA CCGATCGCGT	
TCACCCAGGC CGTCCTTACG GCCACCAACT TCTTCGGTAT CAACACGATC	420
CGATCGCGT GGATTATTC ATCCGTATGT GGAACCAGGC AGCCCTGGCA ATGGAGGTCT	480
ACCAGGCCGA GACCGCGTT AACACGCTT TCGAGAACGCT CGAGCCGATG GCGTCGATCC	540
TGAGTCCCGG CGCGAGCCAG AGCACGACGA ACCCGATCTT CGGAATGCC	600
TCCCCTGGCA GCTCAACACC GGTTGGCCAG TTGCCGCCGG CGGCTACCCA GACCCTCGGC	660
CAACTGGGTG AGATGAGCGG CCCGATGCAG CAGCTGACCC AGCCGCTGCA GCAGGTTGACG	720
TCGTTGTTCA GCCAGGTGGG CGGCACCGGC GGCAGCAACC CAGCCGACGA GGAAGCCGCG	780
CAGATGGGCC TGCTCGGCAC CAGTCGCTG TCGAACCATC CGCTGGCTGG TGGATCAGGC	840
CCCAGCGCGG GCGCGGGCCT GCTGCGCGCG GAGTCGCTAC CTGGCGCAGG TGGGTCGTTG	900
ACCCGCACGC CGCTGATGTC TCAGCTGATC GAAAAGCCGG TTGCCCTTC GGTGATGCCG GCGGCTGCTG	960
CCGGATCGTC GGCGACGGGT GGCGCCGCTC CGGTGGGTGC GGGAGCGATG GGCCAGGGTG	1020

145

CGCAATCCGG CGGCTCCACC AGGCCGGGTC TGGTCGCGCC GGCACCGCTC GCGCAGGAGC	1080
GTGAAGAAGA CGACGAGGAC GACTGGGACG AAGAGGACGA CTGGTGAGCT CCCGTAATGA	1140
CAACAGACTT CCGGCCACC CGGGCCGGAA GACTTGCAA CATTGGCG AGGAAGGTA	1200
AGAGAGAAAG TAGTCCAGCA TGGCAGAGAT GAAGACCGAT GCCGCTACCC TCGCGC	1256

## (2) INFORMATION FOR SEQ ID NO:113:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 432 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:113:

CTAGTGGATG GGACCATGGC CATTTCCTGC AGTCTCACTG CCTTCTGTGT TGACATTTG	60
GCACGCCGGC GGAAACGAAG CACTGGGTC GAAGAACGGC TGCGCTGCCA TATCGTCCGG	120
AGCTTCCATA CCTTCGTGCG GCCGGAAGAG CTTGTCGTAG TCGGCCGCCA TGACAACCTC	180
TCAGAGTGCG CTCAAACGTA TAAACACGAG AAAGGGCGAG ACCGACGGAA GGTCGAACTC	240
GCCCCGATCCC GTGTTTCGCT ATTCTACGCG AACTCGGCGT TGCCCTATGC GAACATCCA	300
GTGACGTTGC CTTCGGTCGA AGCCATTGCC TGACCGGCTT CGCTGATCGT CCGCGCCAGG	360
TTCTGCAGCG CGTTGTTCA GCTCGTAGCC GTGGCGTCCC ATTTTGCTG GACACCCTGG	420
TACGCCCTCCG AA	432

## (2) INFORMATION FOR SEQ ID NO:114:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 368 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:114:

Met Leu Trp His Ala Met Pro Pro Glu Xaa Asn Thr Ala Arg Leu Met  
1 5 10 15

Ala Gly Ala Gly Pro Ala Pro Met Leu Ala Ala Ala Ala Gly Trp Gln  
20 25 30

Thr Leu Ser Ala Ala Leu Asp Ala Gln Ala Val Glu Leu Thr Ala Arg  
35 40 45

Leu Asn Ser Leu Gly Glu Ala Trp Thr Gly Gly Ser Asp Lys Ala  
50 55 60

Leu Ala Ala Ala Thr Pro Met Val Val Trp Leu Gln Thr Ala Ser Thr  
65 70 75 80

Gln Ala Lys Thr Arg Ala Met Gln Ala Thr Ala Gln Ala Ala Ala Tyr  
85 90 95

Thr Gln Ala Met Ala Thr Thr Pro Ser Leu Pro Glu Ile Ala Ala Asn  
100 105 110

His Ile Thr Gln Ala Val Leu Thr Ala Thr Asn Phe Phe Gly Ile Asn  
115 120 125

Thr Ile Pro Ile Ala Leu Thr Glu Met Asp Tyr Phe Ile Arg Met Trp  
130 135 140

Asn Gln Ala Ala Leu Ala Met Glu Val Tyr Gln Ala Glu Thr Ala Val  
145 150 155 160

Asn Thr Leu Phe Glu Lys Leu Glu Pro Met Ala Ser Ile Leu Asp Pro  
165 170 175

Gly Ala Ser Gln Ser Thr Thr Asn Pro Ile Phe Gly Met Pro Ser Pro  
180 185 190

Gly Ser Ser Thr Pro Val Gly Gln Leu Pro Pro Ala Ala Thr Gln Thr  
195 200 205

Leu Gly Gln Leu Gly Glu Met Ser Gly Pro Met Gln Gln Leu Thr Gln  
210 215 220

Pro Leu Gln Gln Val Thr Ser Leu Phe Ser Gln Val Gly Gly Thr Gly  
225 230 235 240

Gly Gly Asn Pro Ala Asp Glu Glu Ala Ala Gln Met Gly Leu Leu Gly  
245 250 255

Thr Ser Pro Leu Ser Asn His Pro Leu Ala Gly Gly Ser Gly Pro Ser  
260 265 270

Ala Gly Ala Gly Leu Leu Arg Ala Glu Ser Leu Pro Gly Ala Gly Gly  
275 280 285

Ser Leu Thr Arg Thr Pro Leu Met Ser Gln Leu Ile Glu Lys Pro Val  
290 295 300

Ala Pro Ser Val Met Pro Ala Ala Ala Gly Ser Ser Ala Thr Gly  
305 310 315 320

Gly Ala Ala Pro Val Gly Ala Gly Ala Met Gly Gln Gly Ala Gln Ser  
325 330 335

Gly Gly Ser Thr Arg Pro Gly Leu Val Ala Pro Ala Pro Leu Ala Gln  
340 345 350

Glu Arg Glu Glu Asp Asp Glu Asp Asp Trp Asp Glu Glu Asp Asp Trp  
355 360 365

## (2) INFORMATION FOR SEQ ID NO:115:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 12 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:115:

Met Ala Glu Met Lys Thr Asp Ala Ala Thr Leu Ala  
1 5 10

## (2) INFORMATION FOR SEQ ID NO:116:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 396 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:116:

GATCTCCGGC GACCTGAAAA CCCAGATCGA CCAGGTGGAG TCGACGGCAG GTTCGTTGCA	60
GGGCCAGTGG CGCGGCGCGG CGGGGACGGC CGCCCAGGCC GCGGTGGTGC GCTTCCAAGA	120
AGCAGCCAAT AAGCAGAACG AGGAACTCGA CGAGATCTCG ACGAATATTG GTCAGGCCGG	180
CGTCCAATAC TCGAGGGCCG ACGAGGAGCA GCAGCAGGCG CTGTCTCGC AAATGGGCTT	240
CTGACCCGCT AATACGAAAA GAAACGGAGC AAAAACATGA CAGAGCAGCA GTGGAATTTC	300
GCAGGTATCG AGGCCGCGGC AAGCGCAATC CAGGGAAATG TCACGTCCAT TCATTCCCTC	360
CTTGACGAGG GGAAGCAGTC CCTGACCAAG CTCGCA	396

## (2) INFORMATION FOR SEQ ID NO:117:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 80 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:117:

Ile Ser Gly Asp Leu Lys Thr Gln Ile Asp Gln Val Glu Ser Thr Ala  
1               5                           10                   15  
  
Gly Ser Leu Gln Gly Gln Trp Arg Gly Ala Ala Gly Thr Ala Ala Gln  
20              25                           30  
  
Ala Ala Val Val Arg Phe Gln Glu Ala Ala Asn Lys Gln Lys Gln Glu  
35              40                           45  
  
Leu Asp Glu Ile Ser Thr Asn Ile Arg Gln Ala Gly Val Gln Tyr Ser  
50              55                           60  
  
Arg Ala Asp Glu Glu Gln Gln Ala Leu Ser Ser Gln Met Gly Phe  
65              70                           75                   80

## (2) INFORMATION FOR SEQ ID NO:118:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 387 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:118:

GTGGATCCCG ATCCCGTGT TCGCTATTCT ACGCGAACTC GGC GTTGCCC TATGCGAAC 60  
TCCCAGTGAC GTTGCCTTCG GTCGAAGCCA TTGCCTGACC GGCTTCGCTG ATCGTCCGCG 120  
CCAGGTTCTG CAGCGCGTTG TTCAGCTCGG TAGCCGTGGC GTCCCATTTC TGCTGGACAC 180  
CCTGGTACGC CTCCGAACCG CTACCGCCCC AGGCCGCTGC GAGCTTGTC AGGGACTGCT 240  
TCCCCTCGTC AAGGAGGGAA TGAATGGACG TGACATTCC CTGGATTGCG CTTGCCGCGG 300  
CCTCGATAACC CGCGAAATTG CACTGCTGCT CTGTCATGTT TTTGCTCCGT TTCTTTCGT 360  
ATTAGCGGGT CAGAAGCCCA TTTGCGA 387

150

## (2) INFORMATION FOR SEQ ID NO:119:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 272 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:119:

CGGCACGAGG ATCTCGGTTG GCCCAACGGC GCTGGCGAGG GCTCCGTTCC GGGGGCGAGC	60
TGCGCGCCGG ATGCTTCCTC TGCCCGCAGC CGCGCCTGGA TGGATGGACC AGTTGCTACC	120
TTCCCGACGT TTCGTTCGGT GTCTGTGCGA TAGCGGTGAC CCCGGCGCGC ACGTCGGGAG	180
TGTTGGGGGG CAGGCCGGGT CGGTGGTTCG GCCGGGGACG CAGACGGTCT GGACGGAACG	240
GGCGGGGGTT CGCCGATTGG CATCTTGCC CA	272

## (2) INFORMATION FOR SEQ ID NO:120:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:120:

Asp Pro Val Asp Ala Val Ile Asn Thr Thr Cys Asn Tyr Gly Gln Val			
1	5	10	15
Val Ala Ala Leu			
20			

## (2) INFORMATION FOR SEQ ID NO:121:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 15 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:121:

Ala Val Glu Ser Gly Met Leu Ala Leu Gly Thr Pro Ala Pro Ser  
1               5                           10                           15

(2) INFORMATION FOR SEQ ID NO:122:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 19 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:122:

Ala Ala Met Lys Pro Arg Thr Gly Asp Gly Pro Leu Glu Ala Ala Lys  
1               5                           10                           15  
  
Glu Gly Arg

(2) INFORMATION FOR SEQ ID NO:123:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 15 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:123:

Tyr Tyr Trp Cys Pro Gly Gln Pro Phe Asp Pro Ala Trp Gly Pro  
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:124:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:124:

Asp Ile Gly Ser Glu Ser Thr Glu Asp Gln Gln Xaa Ala Val  
1 5 10

(2) INFORMATION FOR SEQ ID NO:125:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:125:

Ala Glu Glu Ser Ile Ser Thr Xaa Glu Xaa Ile Val Pro  
1 5 10

(2) INFORMATION FOR SEQ ID NO:126:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:126:

Asp Pro Glu Pro Ala Pro Pro Val Pro Thr Thr Ala Ala Ser Pro Pro  
1               5                           10                           15  
Ser

(2) INFORMATION FOR SEQ ID NO:127:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 15 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:127:

Ala Pro Lys Thr Tyr Xaa Glu Glu Leu Lys Gly Thr Asp Thr Gly  
1               5                           10                           15

(2) INFORMATION FOR SEQ ID NO:128:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 30 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:128:

Asp Pro Ala Ser Ala Pro Asp Val Pro Thr Ala Ala Gln Leu Thr Ser  
1 5 10 15  
Leu Leu Asn Ser Leu Ala Asp Pro Asn Val Ser Phe Ala Asn  
20 25 30

(2) INFORMATION FOR SEQ ID NO:129:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:129:

Asp Pro Pro Asp Pro His Gln Xaa Asp Met Thr Lys Gly Tyr Tyr Pro  
1 5 10 15  
Gly Gly Arg Arg Xaa Phe  
20

(2) INFORMATION FOR SEQ ID NO:130:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:130:

Asp Pro Gly Tyr Thr Pro Gly  
1 5

## (2) INFORMATION FOR SEQ ID NO:131:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 10 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: linear

## (ix) FEATURE:

(D) OTHER INFORMATION: /note= "The Second Residue Can Be Either a Pro or Thr"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:131:

Xaa Xaa Gly Phe Thr Gly Pro Gln Phe Tyr  
1               5                           10

## (2) INFORMATION FOR SEQ ID NO:132:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 9 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: linear

## (ix) FEATURE:

(D) OTHER INFORMATION: /note= "The Third Residue Can Be Either a Gln or Leu"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:132:

Xaa Pro Xaa Val Thr Ala Tyr Ala Gly  
1               5

## (2) INFORMATION FOR SEQ ID NO:133:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 9 amino acids

- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:133:

Xaa Xaa Xaa Glu Lys Pro Phe Leu Arg  
1 5

(2) INFORMATION FOR SEQ ID NO:134:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 15 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:134:

Xaa Asp Ser Glu Lys Ser Ala Thr Ile Lys Val Thr Asp Ala Ser  
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:135:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 15 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:135:

Ala Gly Asp Thr Xaa Ile Tyr Ile Val Gly Asn Leu Thr Ala Asp

1                   5                   10                   15

(2) INFORMATION FOR SEQ ID NO:136:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 15 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:136:

Ala Pro Glu Ser Gly Ala Gly Leu Gly Gly Thr Val Gln Ala Gly  
1                 5                   10                   15

(2) INFORMATION FOR SEQ ID NO:137:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 21 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS:
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:137:

Xaa Tyr Ile Ala Tyr Xaa Thr Thr Ala Gly Ile Val Pro Gly Lys Ile  
1                 5                   10                   15  
  
Asn Val His Leu Val  
20

Claims

1. A polypeptide comprising an immunogenic portion of a soluble *M. tuberculosis* antigen, or a variant of said antigen that differs only in conservative substitutions and/or modifications, wherein said antigen has an N-terminal sequence selected from the group consisting of:

- (a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Thr-Cys-Asn-Tyr-Gly-Gln-Val-Val-Ala-Ala-Leu; (SEQ ID No. 120)
- (b) Ala-Val-Glu-Ser-Gly-Met-Leu-Ala-Leu-Gly-Thr-Pro-Ala-Pro-Ser; (SEQ ID No. 121)
- (c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-Ala-Lys-Glu-Gly-Arg; (SEQ ID No. 122)
- (d) Tyr-Tyr-Trp-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Trp-Gly-Pro; (SEQ ID No. 123)
- (e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val; (SEQ ID No. 124)
- (f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro; (SEQ ID No. 125)
- (g) Asp-Pro-Glu-Pro-Ala-Pro-Pro-Val-Pro-Thr-Thr-Ala-Ala-Ser-Pro-Pro-Ser; (SEQ ID No. 126)
- (h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly; (SEQ ID No. 127)
- (i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Leu-Thr-Ser-Leu-Leu-Asn-Ser-Leu-Ala-Asp-Pro-Asn-Val-Ser-Phe-Ala-Asn; (SEQ ID No. 128) and
- (j) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Leu-Gly-Gly-Thr-Val-Gln-Ala-Gly; (SEQ ID No. 136)

wherein Xaa may be any amino acid.

2. A polypeptide comprising an immunogenic portion of an *M. tuberculosis* antigen, or a variant of said antigen that differs only in conservative

substitutions and/or modifications, wherein said antigen has an N-terminal sequence selected from the group consisting of:

- (a) Asp-Pro-Pro-Asp-Pro-His-Gln-Xaa-Asp-Met-Thr-Lys-Gly-Tyr-Tyr-Pro-Gly-Gly-Arg-Arg-Xaa-Phe; (SEQ ID No. 129) and
- (b) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn-Val-His-Leu-Val; (SEQ ID No. 137), wherein Xaa may be any amino acid.

3. A polypeptide comprising an immunogenic portion of a soluble *M. tuberculosis* antigen, or a variant of said antigen that differs only in conservative substitutions and/or modifications, wherein said antigen comprises an amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID Nos.: 1, 2, 4-10, 13-25, 52, 99 and 101, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID Nos.: 1, 2, 4-10, 13-25, 52, 99 and 101 or a complement thereof under moderately stringent conditions.

4. A polypeptide comprising an immunogenic portion of a *M. tuberculosis* antigen, or a variant of said antigen that differs only in conservative substitutions and/or modifications, wherein said antigen comprises an amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID Nos.: 26-51, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID Nos.: 26-51 or a complement thereof under moderately stringent conditions.

5. A DNA molecule comprising a nucleotide sequence encoding a polypeptide according to any one of claims 1-4.

6. An expression vector comprising a DNA molecule according to claim 5.

7. A host cell transformed with an expression vector according to claim 6.

8. The host cell of claim 7 wherein the host cell is selected from the group consisting of *E. coli*, yeast and mammalian cells.

9. A pharmaceutical composition comprising one or more polypeptides according to any one of claims 1-4 and a physiologically acceptable carrier.

10. A pharmaceutical composition comprising one or more DNA molecules according to claim 5 and a physiologically acceptable carrier.

11. A pharmaceutical composition comprising one or more DNA sequences recited in SEQ ID Nos.: 3, 11 and 12; and a physiologically acceptable carrier.

12. A vaccine comprising one or more polypeptides according to any one of claims 1-4 and a non-specific immune response enhancer.

13. A vaccine comprising:

a polypeptide having an N-terminal sequence selected from the group consisting of sequences recited in SEQ ID NO: 134 and 135; and

a non-specific immune response enhancer.

14. A vaccine comprising:

one or more polypeptides encoded by a DNA sequence selected from the group consisting of SEQ ID Nos.: 3, 11 and 12, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID Nos.: 3, 11 and 12; and

a non-specific immune response enhancer.

15. The vaccine of claims 12-14 wherein the non-specific immune response enhancer is an adjuvant.

16. A vaccine comprising one or more DNA molecules according to claim 5 and a non-specific immune response enhancer.

17. A vaccine comprising one or more DNA sequences recited in SEQ ID Nos.: 3, 11 and 12; and a non-specific immune response enhancer.

18. The vaccine of claims 16 or 17 wherein the non-specific immune response enhancer is an adjuvant.

19. A method for inducing protective immunity in a patient, comprising administering to a patient a pharmaceutical composition according to any one of claims 9-11.

20. A method for inducing protective immunity in a patient, comprising administering to a patient a vaccine according to any one of claims 12-18.

21. A fusion protein comprising two or more polypeptides according to any one of claims 1-4.

22. A fusion protein comprising one or more polypeptides according to any one of claims 1-4 and ESAT-6.

23. A pharmaceutical composition comprising a fusion protein according to claim 21 or 22 and a physiologically acceptable carrier.

24. A vaccine comprising a fusion protein according to claims 21 or 22 and a non-specific immune response enhancer.

25. The vaccine of claim 24 wherein the non-specific immune response enhancer is an adjuvant.

26. A method for inducing protective immunity in a patient, comprising administering to a patient a pharmaceutical composition according to claim 23.

27. A method for inducing protective immunity in a patient, comprising administering to a patient a vaccine according to claims 24 or 25.

28. A method for detecting tuberculosis in a patient, comprising:
  - (a) contacting dermal cells of a patient with one or more polypeptides according to any one of claims 1-4; and
    - (b) detecting an immune response on the patient's skin and therefrom detecting tuberculosis in the patient.
29. A method for detecting tuberculosis in a patient, comprising:
  - (a) contacting dermal cells of a patient with a polypeptide having an N-terminal sequence selected from the group consisting of sequences recited in SEQ ID NO: 134 and 135; and
    - (b) detecting an immune response on the patient's skin and therefrom detecting tuberculosis in the patient.
30. A method for detecting tuberculosis in a patient, comprising:
  - (a) contacting dermal cells of a patient with one or more polypeptides encoded by a DNA sequence selected from the group consisting of SEQ ID Nos.: 3, 11 and 12, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID Nos.: 3, 11 and 12; and
    - (b) detecting an immune response on the patient's skin and therefrom detecting tuberculosis in the patient.
31. The method of any one of claims 28-30 wherein the immune response is induration.
32. A diagnostic kit comprising:
  - (a) a polypeptide according to any one of claims 1-4; and
  - (b) apparatus sufficient to contact said polypeptide with the dermal cells of a patient.

33. A diagnostic kit comprising:

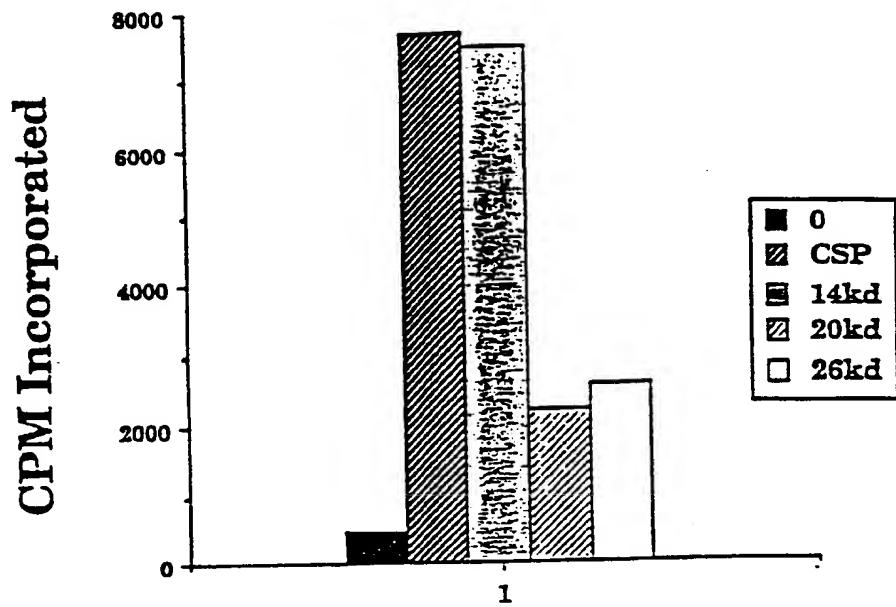
- (a) a polypeptide having an N-terminal sequence selected from the group consisting of sequences recited in SEQ ID NO: 134 and 135; and
- (b) apparatus sufficient to contact said polypeptide with the dermal cells of a patient.

34. A diagnostic kit comprising:

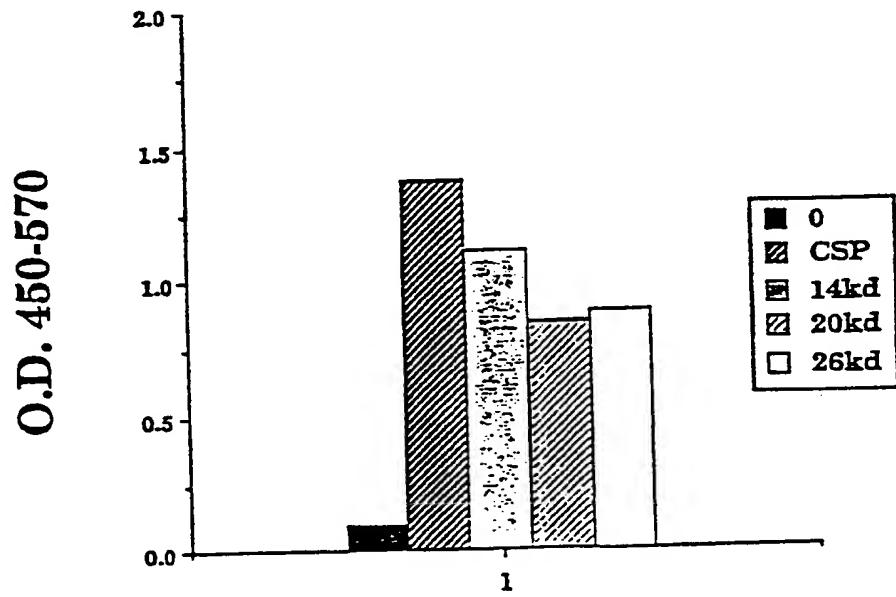
- (a) a polypeptide encoded by a DNA sequence selected from the group consisting of SEQ ID Nos.: 3, 11 and 12, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID Nos.: 3, 11 and 12; and
- (b) apparatus sufficient to contact said polypeptide with the dermal cells of a patient.

1/3

### D7 T Cell Proliferation

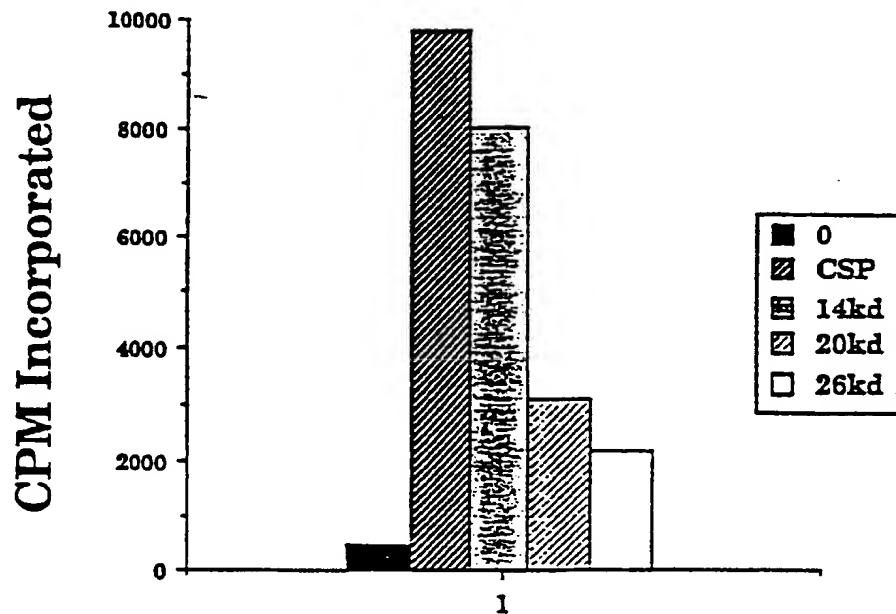
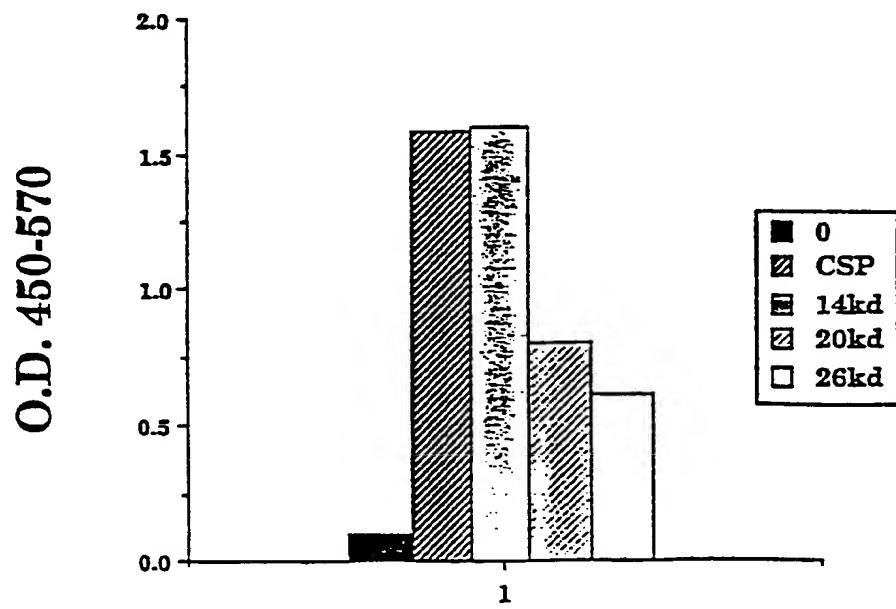


### D7 IFNg



*Fig. 1A*

2/3

**D160 T Cell Proliferation****D160 IFNg***Fig. 1B*

3/3

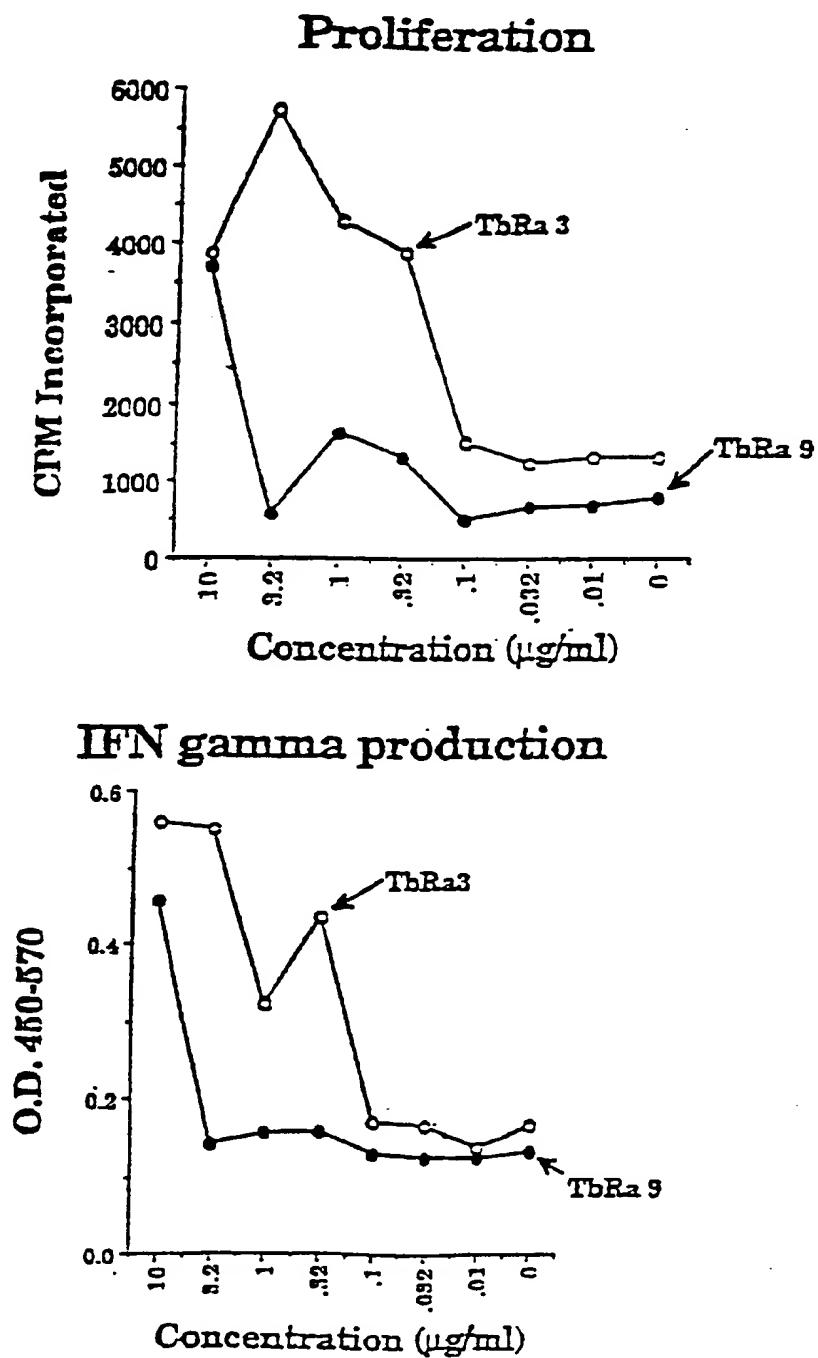


Fig. 2